

Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome

**Mathilde BARTHE^{1*}, Claire DOUTRELANT², Rita COVAS³⁻⁵, Martim MELO³⁻⁵,
Juan Carlos ILLERA⁶, Marie-Ka TILAK¹, Constance COLOMBIER¹, Thibault
LEROY^{1,7}, Claire LOISEAU^{2,3a}, Benoit NABHOLZ^{1,8a}**

¹ ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France

² Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, Univ Montpellier, EPHE, IRD, Montpellier, France

³ CIBIO, Research Centre in Biodiversity and Genetic Resources, InBIO Associated Laboratory, Campus Agrário de Vairão, Vairão, Portugal

⁴ MHNC-UP, Natural History and Science Museum of the University of Porto, Porto, Portugal

⁵ DST/NRF Centre of Excellence, FitzPatrick Institute, University of Cape Town, Rondebosch, South Africa

⁶ Biodiversity Research Institute (CSIC-Oviedo University-Principality of Asturias), Oviedo University, Mieres, Spain

⁷IRHS-UMR1345, Université d'Angers, INRAE, Institut Agro, SFR 4207 QuaSaV, 49071, Beaucouzé, France

⁸Institut universitaire de France, Paris

* Corresponding author : mathilde.barthe.pro@gmail.com

^a Last co-authors

1 **Abstract**

2 Shared ecological conditions encountered by species that colonize islands often lead to the
3 evolution of convergent phenotypes, commonly referred to as [the](#) “island syndrome”. Reduced
4 immune functions have been previously proposed to be part of [this](#) syndrome, as a consequence
5 of the reduced diversity of pathogens on island ecosystems. According to this hypothesis,
6 immune genes are expected to exhibit genomic signatures of relaxed selection pressure in
7 island species. In this study, we used comparative genomic methods to study immune genes in
8 island species (N = 20) and their mainland relatives (N = 14). We gathered public data as well
9 as generated new data on innate ([TLR](#): Toll-Like Receptors, [BD](#): Beta Defensins) and acquired
10 immune genes ([MHC](#): Major Histocompatibility Complex classes I and II), but also on
11 hundreds of genes with various immune functions. As a control, we used a set of 97 genes, [not](#)
12 [known as](#) involved in immune functions [based on the literature](#), to account [the increased drift](#)
13 [effects](#) for the lower effective population sizes in island species. We used synonymous and
14 non-synonymous variations to estimate the selection pressure acting on immune genes. [BDs](#)
15 [and TLRs have higher ratios of non-synonymous over synonymous polymorphisms \(Pn/Ps\)](#)
16 [than randomly selected control genes suggesting that they evolve under a different selection](#)
17 [regime than non-immune related genes. However, simulations analyses show that this is](#)
18 [unlikely to be explained by ongoing positive selection or balancing selection.](#) For the MHC
19 evolving under balancing selection, we used simulation to estimate the impact of population
20 size variation. We found a significant effect of drift on immune genes of island species leading
21 to a reduction in genetic diversity and efficacy of selection. However, the intensity of relaxed
22 selection was not significantly different from control genes, except for MHC class II genes.
23 These genes exhibit a significantly higher level of non-synonymous loss of polymorphism than
24 expected assuming only drift and an evolution under frequency dependent selection, possibly
25 due to a reduction of extracellular parasite communities on islands. Overall, our results showed
26 that demographic effects lead to a decrease in the immune functions of island species, but the
27 relaxed selection caused by a reduced parasite pressure may only occur in some immune genes
28 categories.

29 **Keywords:** genetic drift, island evolution, immunity, Toll-Like Receptors, Beta-Defensins,
30 major histocompatibility complex, molecular evolution, population genomics

32 **Introduction**

33 Island colonizers face new communities of competitors, predators and parasites in a small area
34 with limited resources, which generally result in high extinction rates of colonizers (Losos and
35 Ricklefs, 2009). Oceanic island faunas are characterized by a low species richness, coupled
36 with high population densities for each species (MacArthur and Wilson, 1967; Warren et al.,
37 2015) - which translates in communities with, on average, low levels of inter-specific
38 interaction and high levels of intra-specific competition (but see Rando et al., 2010 for an
39 example of character displacement due to competition among island finch species). These
40 shared island characteristics are thought to underlie the evolution of convergent phenotypes, in
41 what is called the ‘island syndrome’ (Baeckens and Van Damme, 2020). Convergence has been
42 documented in multiple traits, such as size modification (dwarfism or gigantism; Lomolino,
43 2005), reduction of dispersal (Baeckens and Van Damme, 2020) shift towards K life history
44 strategies (Boyce, 1984; Covas, 2012; MacArthur and Wilson, 1967), evolution of generalist
45 traits (Blondel, 2000; Warren et al., 2015), or changes in colour and acoustic signals
46 (Doutrelant et al., 2016; Grant, 1965).

47 Reduced immune function has also been hypothesized as an island syndrome trait, directly
48 linked to reduced parasite pressure on islands (Lobato et al., 2017; Matson and Beadell, 2010;
49 Wikelski et al., 2004). Island parasite communities are i) less diverse (Beadell et al., 2006;
50 Illera et al., 2015; Loiseau et al., 2017; Maria et al., 2009; Pérez-Rodríguez et al., 2013), and
51 ii) could be less virulent due to the expansion of the ecological niche expected by the theory of
52 island biogeography. In fact, island parasites are generally more generalists than their mainland
53 counterparts, which could lead to a reduced virulence due to the trade-off between replication
54 capacity and resistance against host immune defenses (Garamszegi, 2006; Hochberg and
55 Møller, 2001; Pérez-Rodríguez et al., 2013). Overall, a reduction of parasitic pressure should
56 lead to a weakening of the immune system due to the costs of maintaining efficient immune
57 functions (Lindström et al., 2004; Matson and Beadell, 2010; Wikelski et al., 2004). Such
58 reduction may have important implications for the ability of these populations to resist or
59 tolerate novel pathogens. The introduction of avian malaria in the Hawaiian archipelago, and
60 the subsequent extinctions and population declines of many endemic species is the most
61 emblematic example (Van Riper III et al., 1986; Wikelski et al., 2004).

62 Immunological parameters, such as blood leukocyte concentration, antibodies or other immune
63 proteins (e.g. haptoglobin), hemolysis, and hemagglutination (Lee et al., 2006; Matson and

64 Beadell, 2010) may serve as proxies to determine population immune functions. To date, the
65 majority of studies that focused on island avifauna have found ambiguous results, with either
66 no support for a reduced immune response on island species (Beadell et al., 2007; Matson,
67 2006), or contrasted results, such as a lower humoral component (total immunoglobulins) on
68 islands, but a similar innate component (haptoglobin levels) between island and mainland
69 species (Lobato et al., 2017). The use of immune parameters as proxies of immune function is
70 fraught with difficulties (Lobato et al., 2017). The study of molecular evolution of immune
71 genes therefore represents an alternative strategy to tackle this question. However, it is
72 necessary to distinguish neutral effects, the demographic effects resulting from island
73 colonization, from selective ones, the potential relaxation of selection pressures due to the
74 changes in the pathogen community.

75 First, the bottleneck experienced by species during island colonization leads to a decrease in
76 genetic variability (Frankham, 1997). A reduced genetic diversity at loci involved in immunity
77 should have a direct implication on immune functions (Hale and Briskie, 2007 but see ; Hawley
78 et al., 2005; Spurgin et al., 2011). Second, small population sizes increase genetic drift, which
79 may counteract the effect of natural selection [on weakly deleterious mutations](#) (Ohta, 1992).
80 Several recent studies found a greater load of deleterious mutations in island species (Kutschera
81 et al., 2020; Leroy et al., 2021b; Loire et al., 2013; Robinson et al., 2016; Rogers and Slatkin,
82 2017). Finally, it is necessary to differentiate genes involved in the innate versus the acquired
83 immune response. The innate immune response is the first line of defense and is composed of
84 phagocytes, macrophages and dendritic cells. These cells allow non-specific recognition of
85 pathogens (Akira, 2003; Alberts et al., 2002). For example, Toll-Like Receptors (TLR;
86 transmembrane proteins) trigger a chain reaction leading to the production of various
87 substances, including antimicrobial peptides such as beta-defensins (BD) that have active
88 properties in pathogen cell lysis (Velová et al., 2018). On the other hand, the acquired immune
89 system allows a specific response, characterized by immune memory. Major
90 Histocompatibility Complex (MHC) genes code for surface glycoproteins that bind to antigenic
91 peptides, and present them to the cells of the immune system; class I and II genes ensure the
92 presentation of a broad spectrum of intra- and extracellular-derived peptides, respectively
93 (Klein, 1986). Although all these genes are directly involved in the identification and
94 neutralization of pathogens, previous studies found that they evolve under different selection
95 regimes: TLRs and BDs are under purifying selection which usually results in the selective

96 removal of deleterious alleles and stabilizing selection (Grueber et al., 2014; van Dijk et al.,
97 2008), whereas MHC genes are under balancing selection (Bernatchez and Landry, 2003).

98 Recent studies on birds (Gonzalez-Quevedo et al., 2015a, 2015b), amphibians (Belasen et al.,
99 2019), and lizards (Santonastaso et al., 2017) found that the demographic history of island
100 populations led to the loss of genetic variation at immune genes involved in pathogen
101 recognition, such as TLRs and MHC. For example, (Santonastaso et al., 2017) demonstrated
102 that the polymorphism pattern in MHC genes and microsatellites covary positively with island
103 area in *Podarcis* lizards, suggesting a dominant role for genetic drift in driving the evolution
104 of the MHC. Gonzalez-Quevedo, et al. (2015a) found a similar pattern comparing TLR and
105 microsatellite polymorphism in the Berthelot pipit, *Anthus berthelotii*, an endemic species from
106 *Macaronesia*, supporting a predominant role of genetic drift in TLR evolution. However, these
107 studies did not explicitly test the hypothesis of a relaxed selection pressure on islands imposed
108 by an impoverished parasite community. All other things being equal, it is expected that the
109 polymorphism pattern of a coding sequence decreases with population size (Buffalo, 2021;
110 Leroy et al., 2021b). Therefore, a decrease in polymorphism with population size could not be
111 taken as a proof of a relaxation in the selection pressure.

112 Here, we study a dataset of 34 bird species (20 insular and 14 mainland species; Figure 1)
113 combining the 24 species of Leroy et al. (2021b) and 10 newly generated by targeted-capture
114 sequencing (Table 1). To be able to demonstrate a change in natural selection, a traditional
115 approach is to contrast polymorphism of synonymous sites (P_s) with polymorphism of non-
116 synonymous sites (P_n). Synonymous mutations refer to mutations that do not alter amino acid
117 sequences, whereas non-synonymous mutations do.

118 Following population genetic theory, in a diploid population, $P_s = 4 N_e \mu$ and $P_n = 4 N_e \mu f$,
119 where N_e is the effective population size, μ is the mutation rate and f is a function that integrates
120 the probability of an allele to segregate at a given frequency. f depends on the distribution of
121 the fitness effect (DFE) of mutations (Eyre-Walker and Keightley, 2007). This distribution
122 scales with N_e as the fitness effect is dependent on N_e multiplied by the coefficient of selection
123 s (Kimura, 1962). The nearly-neutral theory predicts that the DFE includes a large proportion
124 of mutations with a $N_e * s$ close to 0 (Ohta, 1992). As a consequence, an increase of N_e will lead
125 to an increase of the fitness effect of weakly deleterious mutations, in such a way that these
126 mutations will be more easily removed from the population by natural selection, therefore
127 reducing P_n relative to P_s leading to a negative correlation between P_n/P_s and P_s (through N_e ;

128 Welch et al., 2008). The presence of linked positively selected mutations does not change
129 qualitatively this relationship (Castellano et al., 2018; Chen et al., 2020 and our simulations
130 below).

131 Shifts in the parasitic community on islands are expected to have an impact on the ratio Pn/Ps
132 of immune genes. However, the fixation probability depends on the product $Ne s$, and a
133 variation in Ne is also expected to impact the efficacy of selection and thus the ratio Pn/Ps
134 across the entire transcriptome, particularly in the presence of slightly deleterious mutations
135 (Charlesworth and Eyre-Walker, 2008; Leroy et al., 2021b; Loire et al., 2013; Ohta, 1992). In
136 addition, due to their lower population sizes, island birds compared to continental species
137 exhibit a genome-wide reduction in genetic diversity and efficacy of selection (Kutschera et
138 al., 2020; Leroy et al., 2021b). Therefore, we expect a similar reduction in immune genes
139 diversity even without any change in the parasite pressure.

140 To disentangle the effect of population size from a change in parasite pressure and estimate the
141 impact of demography on the efficacy of selection, we randomly selected protein-coding genes
142 (i.e., control genes) involved in various biological functions (Fijarczyk et al., 2016; Leroy et
143 al., 2021b). The selection pressure acting on the randomly selected control genes is expected
144 to be similar between island and mainland bird species. Therefore, the variation of Pn/Ps of the
145 control genes is only dependent on the variation of Ne . In contrast, if a reduced parasite pressure
146 on islands directly impacts the evolution of immune genes, the Pn/Ps of immune genes is
147 expected to show a larger variation between island and continental species than the control
148 genes. More specifically, for genes under purifying selection, non-synonymous weakly
149 deleterious mutations, normally eliminated under strong selection, would be maintained,
150 leading to an increase of Pn/Ps. By contrast, for genes under balancing selection, non-
151 synonymous advantageous mutations, normally maintained in the polymorphism under strong
152 selection, would be fixed or eliminated leading to a decrease of Pn/Ps (Figure 2).

153

154

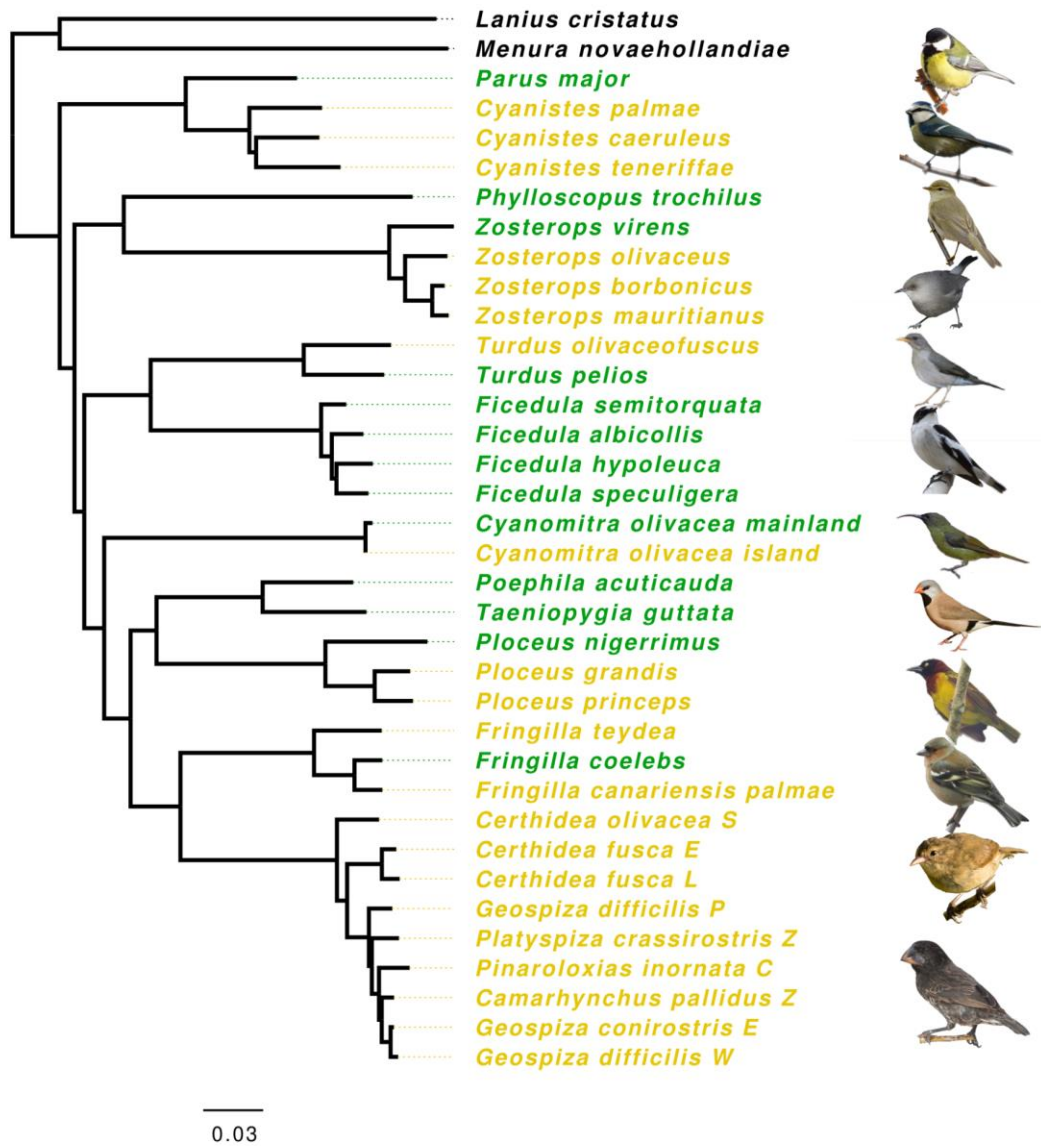
155

156

157

158

159



160

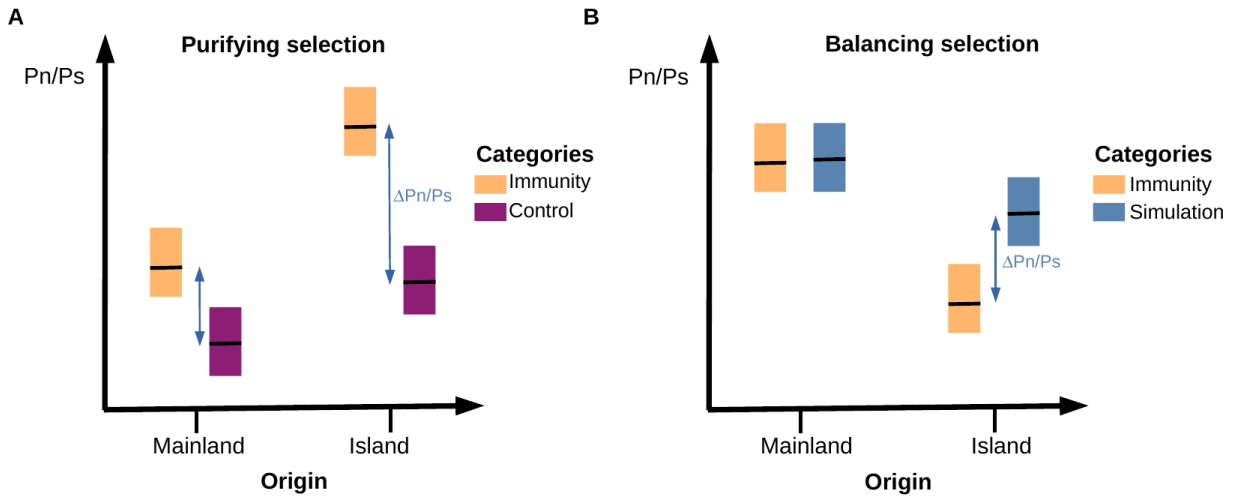
161 **Figure 1:** Phylogeny based on mitochondrial genes of species from the dataset reconstructed by
 162 maximum likelihood method (IQTREE model GTR+Gamma). Species names in yellow indicate island
 163 species, and in green, mainland species. Ultrafast bootstrap values are provided in the supplementary
 164 methods. Some relationships are poorly supported. Bird representations are not to scale. Photos from
 165 top to bottom : *P. major*, *C. caeruleus*, *P. trochilus*, *Z. borbonicus*, *T. pelios*, *F. albicollis*, *C. olivacea*,
 166 *P. acuticauda*, *P. grandis*, *F. coelebs*, *C. fusca*, *G. conirostris*. Photo credits: A. Chudý, F. Desmoulins,
 167 E. Giacone, G. Lasley, Lianaj, Y. Lyubchenko, B. Nabholz, J.D. Reynolds, K. Samodurov, A.
 168 Sarkisyan, Wimvz, Birdpics, T. Aronson, G. Lasley, P. Vos (iNaturalist.org); M. Gabrielli (*Zosterops*
 169 *borbonicus*).

170

171

172

173



174

175 Figure 2: Conceptual diagram showing the expected results under the hypothesis of a relaxation in the
176 selection pressure of the **immune** genes in island species due to a change in the parasitic community.
177 A) Genes evolving under purifying selection where control genes are randomly selected protein-coding
178 genes. B) Genes evolving under balancing selection where controls are obtained from SLiM simulations
179 of genes evolving under the same balancing selection but different population size. Under the hypothesis
180 of a relaxed selection as a consequence of the reduced diversity of pathogens on island ecosystems, the
181 difference in Pn/Ps between categories ($\Delta Pn/Ps$) is expected to be different between species' origin,
182 leading to a statistical interaction between gene categories and origin.

183

184 **Methods**

185 *Dataset*

186 Alignments of Coding DNA Sequences (CDS) of individuals from 24 species were obtained
187 from Leroy et al. (2021b). In addition, data for ten other species (six and four from islands and
188 mainland, respectively) were newly generated for this study by targeted-capture sequencing.
189 Blood samples and subsequent DNA extractions were performed by different research teams.
190 The complete dataset consisted of 34 bird species (20 and 14 insular and mainland species
191 respectively; Table 1; Figure 1). We filtered alignments in order to retain only files containing
192 a minimum of five diploid individuals per site (Table 1).

193 Sequence enrichment was performed using MYBaits Custom Target Capture Kit targeting 21
194 immune genes: 10 Toll-Like receptors (TLR), 9 Beta Defensins (BD), 2 Major

195 Histocompatibility Complex (MHC) and 97 control genes (see below). We followed the
 196 manufacturer's protocol (Rohland and Reich, 2012). Illumina high-throughput sequencing
 197 using a paired-end 150 bp strategy was performed by Novogene (Cambridge, UK).

198

199 **Table 1:** List of species and sampling localities, along with the type of data obtained and the
 200 number of individuals (N).

Species	Origin	Island/Country	N	Reference genome	Reference for population genomics data	Type of data
<i>Cyanistes teneriffae palmae</i>	Island	La Palma	15	<i>Cyanistes caeruleus</i> (This study)	(Mueller et al., 2016)	Capture
<i>Cyanistes teneriffae teneriffae</i>	Island	Tenerife	14			
<i>Cyanistes caeruleus</i>	Mainland	France	15			
<i>Parus major</i>	Mainland	Europe	10	<i>Parus major</i> (Laine et al., 2016)	(Corcoran et al., 2017)	Whole genome
<i>Phylloscopus trochilus</i>	Mainland	Europe	9	<i>Phylloscopus trochilus</i> (Lundberg et al., 2017)	(Lundberg et al., 2017)	Whole genome
<i>Zosterops virens</i>	Mainland	South Africa	7	<i>Zosterops borbonicus</i> (Leroy et al., 2021a)	(Leroy et al., 2021b)	Whole genome
<i>Zosterops olivaceus</i>	Island	Réunion	15			
<i>Zosterops mauritanus</i>	Island	Mauritius	9			
<i>Zosterops borbonicus</i>	Island	Réunion	25			
<i>Ficedula semitorquata</i>	Mainland	Europe	20	<i>Ficedula albicollis</i> (Ellegren et al., 2012)	(Ellegren et al., 2012)	Whole genome
<i>Ficedula albicollis</i>	Mainland	Europe	20			
<i>Ficedula speculigera</i>	Mainland	Nord Africa	20			
<i>Ficedula hypoleuca</i>	Mainland	Europe	20			
<i>Turdus olivaceofuscus</i>	Island	São Tomé	15	<i>Turdus pelios</i> (This study)	This study	Capture
<i>Turdus pelios</i>	Mainland	Gabon	15			
<i>Cyanomitra olivacea</i>	Island	Príncipe	15	<i>Cyanomitra olivacea</i> (This study)	This study	Capture
<i>Cyanomitra olivacea</i>	Mainland	Gabon	15			
<i>Ploceus grandis</i>	Island	São Tomé	13	<i>Ploceus cucullatus</i> (This study)	This study	Capture
<i>Ploceus princeps</i>	Island	Príncipe	13			
<i>Ploceus nigerrimus</i>	Mainland	Cameroon Gabon	14			
<i>Poephila acuticauda acuticauda</i>	Mainland	Australia	10	<i>Taeniopygia guttata</i> (Warren et al., 2010)	(Singhal et al., 2015)	Whole genome
<i>Taeniopygia guttata castanotis</i>	Mainland	Australia	19			
<i>Fringilla teydea</i>	Island	Tenerife	10	<i>Fringilla coelebs</i> (Recuerda et al., 2021)	(Leroy et al., 2021b)	Whole genome
<i>Fringilla canariensis palmae</i>	Island	La Palma	15			
<i>Fringilla coelebs</i>	Mainland	Spain	9			
<i>Certhidea olivacea</i>	Island	Santiago (Galápagos)	5	<i>Geospiza fortis</i> (Zhang et al., 2012)	(Lamichhaney et al., 2015)	Whole genome
<i>Certhidea fusca</i>	Island	San Cristobal (Galápagos)	10			
<i>Certhidea fusca</i>	Island	Española (Galápagos)	10			
<i>Geospiza difficilis</i>	Island	Pinta (Galápagos)	10			
<i>Platyspiza crassirostris</i>	Island	Santa Cruz (Galápagos)	5			
<i>Pinaroloxias inornata</i>	Island	Coco (Galápagos)	8			
<i>Camarhynchus pallidus</i>	Island	Santa Cruz (Galápagos)	5			
<i>Geospiza difficilis</i>	Island	Wolf (Galápagos)	8			
<i>Geospiza conirostris</i>	Island	Española (Galápagos)	10			

201

202

203 *Newly generated draft genome sequence*

204 We generated whole genome sequences at moderate coverage (~40X) for *Turdus pelios*,
205 *Ploceus cucullatus* and *Cyanomitra olivacea* (from Gabon). Library preparation from blood
206 DNA samples and Illumina high-throughput sequencing using a paired-end 150 bp strategy
207 were performed at Novogene (Cambridge, UK). Raw reads were cleaned using FastP (vers.
208 0.20.0; Chen et al., 2018). Genomes assemblies were performed using SOAPdenovo (vers.
209 2.04) and Gapcloser (v1.10) (Luo et al., 2012) with parameters “-d 1 -D 2” and a kmers size of
210 33. Protein annotation was performed by homology detection using genBlastG (She et al.,
211 2011; <http://genome.sfu.ca/genblast/download.html>) and the transcriptome of the collared
212 flycatcher (*Ficedula albicollis*; assembly FicAlb1.5; Ellegren et al., 2012) as reference.

213 *Capture data processing*

214 Reads from targeted-capture sequencing were cleaned with FastP (vers. 0.20.0; Chen et al.,
215 2018). Reads of each individual were mapped respectively to the nearest available reference
216 genomes using bwa mem (vers. 0.7.17; Li, 2013; Table 1), with default parameters. Samtools
217 (vers. 1.3.1; Li et al., 2009) and Picard (vers. 1.4.2; Picard Toolkit 2019) were used to convert
218 the mapping files, order and index reads according to their position on the chromosomes (or
219 scaffolds) of the reference genomes or on the draft genomes generated in this study for *Ploceus*,
220 *Cyanomitra* and *Turdus*. Duplicate reads were marked using MarkDuplicates (vers. 1.140;
221 Picard Toolkit 2019). SNP calling was performed with Freebayes (vers. 1.3.1; Garrison and
222 Marth, 2012). Freebayes output file (VCF file) was converted to a fasta file by filtering out
223 sites with a minimum quality of 40 and a sequencing depth between 10 and 1000X (sites outside
224 these thresholds were treated as missing data, i.e., ‘N’). CDS were then extracted from the
225 alignments using the coordinates of the annotations (gff files). CDS were aligned using
226 MACSE (vers. 2.03; Ranwez et al., 2011) to prevent frameshift mutation errors and GNU-
227 parallel (Tange, 2018) was used to parallelise the computation.

228 *Selection and identification of immune and control genes*

229 We defined several groups of immune genes to compare with the control genes. The control
230 group consisted of 97 protein-coding genes randomly selected in the genome of *Zosterops*
231 *borbonicus* (Leroy et al., 2021a). These control genes allowed the estimation of the average
232 selection pressure that a gene, not involved in the immune response, undergoes in the genome.

233 These genes are single copy (absence of paralogue) and have a variable GC content
234 representative of the whole transcriptome.

235 For the immune genes, we selected three sets of genes from i) a limited set of genes (Core
236 Group) where functions are unambiguously related to immunity, and ii) two larger sets of genes
237 (Database-group & Sma3s-group), obtained through an automatic annotation pipeline.

238 The Core Group included MHC class I and class II genes, 10 Toll-Like Receptors (TLRs;
239 Velová et al., 2018) and 9 Beta Defensins (BD; Chapman et al., 2016). The Database group
240 included genes identified by Immunome Knowledge Base (Ortutay and Vihinen, 2009,
241 <http://structure.bmc.lu.se/idbase/IKB/>; last access 04/02/2020) and InnateDB (Breuer et al.,
242 2013, <http://www.innatedb.com>; last access 04/02/2020). We also added a set of genes for
243 which the genetic ontology indicated a role in immune functions. To do so, we used the chicken
244 (*Gallus gallus*) annotation (assembly GRCg6a downloaded from Ensembl database in March
245 2020; <https://www.ensembl.org/>). We identified genes with the terms "immun*" or
246 "pathogen*" in their Gene Ontology identifiers description (directory obtained from
247 <http://geneontology.org/>). This set included 2605 genes considered to be involved in immunity,
248 although some may be only indirectly involved in immunity or have a small impact on immune
249 functions. Finally, the third set of genes (Sma3s-group) has been built up through the Sma3s-
250 group program (vers. 2; Munoz-Mérida et al., 2014). This program annotated sequences in
251 order to be associated with biological functions through gene ontology identifiers. The
252 annotation of the genome of *F. albicollis* allowed us to identify 3136 genes associated with the
253 genetic ontology "immune system processes". Like for the Database group, this set may include
254 genes with various functions in the immune response. It should be noted that Sma3s-group and
255 Database-group are not mutually exclusive, and some genes are present in both groups. An
256 analysis was performed to identify and exclude genes under balancing selection from Database-
257 group and Sma3s-group sets using BetaScan (vers. 2; Siewert and Voight, 2020), due to the
258 potentially antagonistic responses of these genes. **Very few genes (only 2 and 3 genes from
259 Database-group and Sma3s-group sets) were identified and removed from the analysis** (see
260 Detection of genes under balancing selection in Supplementary Methods).

261 *Test for contamination and population structure*

262 We use the program CroCo (vers. 1.1; Simion et al., 2018) to identify candidates for cross-
263 species contamination (see supplementary materials for details). Overall, we did not detect a

264 clear case of cross-species contamination in our dataset (Figure S1). Contigs identified as
265 potential contamination always involve a pair of species belonging to the same genus. In this
266 case, contamination could be difficult to identify due to the low genetic divergence between
267 species.

268 For the newly sequences species, we also performed PCA analyses on using allele frequencies
269 of control genes. We use the function `dudi.pca` of `ade4` R packages (Jombart and Ahmed,
270 2011). This analysis aims to check for population structure and to detect potentially
271 problematic individuals (i.e., contaminated individuals). This analysis led to the exclusion of 4
272 individuals (*Ploceus princeps* P6-174; *P. grandis* ST10_094; *P. nigerrimus* G3_016; *C.*
273 *teneriffae* TF57) for which we suspected contamination. Otherwise, no extra population
274 structure was detected (Figure S2-S4).

275 *Hidden paralogy*

276 We compute the statistic $F_{is} = 1 - H_0/H_e$ where H_0 is the average number of heterozygous
277 individual observed ($H_0 = \#heterozygous / n$; where n is the sample size) and H_e is the
278 expected number of heterozygous individuals at Hardy-Weinberg (HW) equilibrium ($H_e =$
279 $(n/(n-1) \sum p_i^2)^{-1}$ where n is the sample size and p_i the allele frequency of a randomly
280 chosen allele). F_{is} varies between -1 and 1 with positive value representing excess of
281 homozygous individuals and negative value representing excess of heterozygous individuals
282 compared to the HW proportions. Gene with high value of nucleotide diversity (P_i) and
283 negative value of F_{is} could represent potential cases where hidden paralogous sequences have
284 not been separated and where all the individuals present heterozygous sites in the positions
285 where a substitution occurred between the paralogous copies. Five TLR21 genes appear
286 problematic ($P_i > 0.01$ and $F_{is} < -0.5$; Figure S5) and were excluded from further analyses.

287 The MHC genes are more difficult to analyse. Indeed, heterozygosity could be comparable to
288 divergence under balancing selection. This makes the identification of orthologs very difficult.
289 We identify a variable number of genes among species (from 1 to 10 genes for MHC class I
290 and MHC class II). We checked the sequence similarity for the 10 copies of the MHC class II
291 in *F. albicollis* and the 7 copies of the MHC class I genes in *C. caeruleus* using `cd-hit` (Fu et
292 al., 2012). For MHC class II, sequence divergences are always higher than 15% indicating that
293 reads will likely be correctly assigned to their corresponding gene copy. For MHC class I,
294 sequence identities could be as high as 95%. In this case, we rely on the fact that the reads from

295 very similar paralogous copies will not be confidently assigned to a gene copy sequence by the
296 mapping software. This will lead to a low mapping score quality and are likely to be discarded
297 during the genotype calling procedure. For example, 3 out of 7 genes of the *Cyanistes* MHC
298 class I genes could not have been correctly genotyped and are missing from our final dataset.

299 ***Data Analysis***

300 *SLiM simulations*

301 We use SLiM (vers. 3.3.2; Haller and Messer, 2017) to estimate the impact of demographic
302 changes on polymorphism patterns under various selection regimes. The following parameters
303 were used in all simulations. Sequences of 30kb with a mutation of $4.6e^{-9}$
304 substitutions/site/generation were simulated (Smeds et al., 2016). Recombination was set to be
305 equal to mutation rate. Introns/exons pattern was reproduced by simulating fragments of 3kb
306 separated by one bp with a very high recombination rate of 0.1 rec./site/generation. Five types
307 of mutations were possible: i) neutral synonymous mutations, ii) non-synonymous mutations
308 with a Distribution of Fitness Effect (DFE) following a gamma law of mean = -0.025 and shape
309 = 0.3, which corresponds to the DFE estimated in Passerines by Rousselle et al. (2020), iii)
310 codominant non-synonymous mutations positively selected with $s = 0.1$, iv) non-synonymous
311 mutations under balancing selection with an effect on fitness initially set at 0.01 but re-
312 estimated by the program at each generation according to the mutation frequency in the
313 population, thus including a frequency-dependent effect and v) non-synonymous mutations
314 under overdominance using with a dominance coefficient of 1.2.

315 We simulate a coding sequence organization where positions one and two of the codons were
316 considered as non-degenerated sites, with the non-synonymous types of mutations previously
317 described are possible in various proportions. The third position was considered as completely
318 neutral where only synonymous mutations could appear.

319 In the absence of control genes evolving under balancing selection, we use SLiM to generate a
320 set of control genes for this category. We simulate two populations of 270,000 and 110,000
321 individuals, representing mainland and island effective population size respectively.

322 We also explore the effect of positive and balancing selection on the pattern of P_s and P_n/P_s in
323 a population of size 50,000, 110,000, 270,000 and 500,000. In order to speed up the
324 computational time, we reduce the population size by a factor 100 and rescale mutation rate,
325 recombination rate and selection coefficient accordingly.

326 All the details of the simulation parameters, calculations of non-synonymous polymorphism
327 rate (Pn) and synonymous polymorphism rate (Ps) of simulated sequences, as well as SLiM
328 command lines are provided in Supplementary Methods and Supplementary Materials.

329 *Polymorphism analyses*

330 Synonymous (Ps) and non-synonymous (Pn) nucleotide diversity were estimated from
331 seq_stat_coding written from the Bio++ library (Available as Supplementary data; Guéguen et
332 al., 2013). The mean Pn/Ps was computed as the sum of Pn over the sum of Ps (Wolf et al.,
333 2009). Ps of concatenated sequences of control genes were estimated for each species of our
334 dataset. For the whole-genome sequence species, we compared the Pn/Ps and Ps estimated
335 obtained using the 97 control genes with the values from Leroy et al., (2021b; ~5000 genes
336 used in their study). Pn/Ps and Ps correlations showed a R² of 0.6 and 0.95 respectively (Figure
337 S6). Thus, the 97 control genes used in our study were representative of a larger set. This
338 allowed us to identify *Phylloscopus trochilus* as an outlier. Unlike for all other species (e.g.
339 *Fringilla coelebs*, Figure S7), **synonymous polymorphism level was correlated to the amount**
340 **of missing data in *P. trochilus* alignments (Figure S7)**. As such, we excluded *P. trochilus* from
341 further analysis.

342 The mean Pn/Ps, calculated from the concatenated sequences of genes from the same gene
343 class (control genes; BD; TLR; MHC I; MHC II; Database-group; Sma3s-group), was
344 estimated for each bird species. Alternative transcripts were identified based on the genomic
345 position in the GFF file. If several transcripts were available, one transcript was randomly
346 selected. Pn/Ps estimates based on less than four polymorphic sites were excluded from the
347 analysis, as were those with no polymorphic non-synonymous sites.

348 *Statistical analyses*

349 To estimate the impact of demographic history on genome-wide polymorphism of island
350 species and the potentially reduced constraints on their immune genes, we computed the ratio
351 of non-synonymous nucleotide diversity over synonymous **nucleotide** diversity (Pn/Ps). A
352 linear mixed model was performed, using the Pn/Ps ratio as dependent variable and, as
353 explanatory variables, the mainland or insular origin of species as well as the category of genes
354 (packages lme4 and lmerTest (Bates et al., 2012; Kuznetsova et al., 2017)). In order to take the
355 phylogenetic effect into account, the taxonomic rank “family” was included as a random effect
356 in the model. **We also used a generalized linear mixed model (using the function glmer of the**

357 package lme4) with the family “Gamma(link=log)” which leads to the same results (Figure
358 S15 to S24). Five linear mixed models were defined i) null model, ii) model with only the
359 origin parameter, iii) model with only the gene category parameter, iv) model using both origin
360 and gene category parameters, and finally v) model including those two parameters and the
361 interaction effect. In some cases, the phylogenetic effect was difficult to estimate because the
362 number of species per family was reduced to one. In that case, we chose to reduce the number
363 of families by grouping Turdidae with Muscicapidae, Nectariniidae, and Estrildidae with
364 Ploceidae and Fringillidae within Thraupidae. The results obtained with these family groupings
365 were similar to the original model (Table S1), except when stated. The categories Database-
366 group and Sma3s-group were tested separately from the Core group because they contained
367 hundreds of genes annotated using the automatic pipeline that were only available for species
368 with genome wide data. Database-group and Sma3s-group were not analysed simultaneously
369 because they contained a partially overlapping set of genes. Finally, genes evolving under
370 purifying selection and genes evolving under balancing selection were also analysed
371 separately. Model selection was based on two methods. First, we use the difference in corrected
372 Akaike Information Criterion ($\Delta AICc$) calculated using the qpcR package (Spiess and Spiess,
373 2018). Second, a model simplification using an ANOVA between models was also performed.

374 We also tested an alternative model using the difference between Pn/Ps of immune genes and
375 control genes ($\Delta Pn/Ps$) as dependent variable, and species origin as explanatory variable. Under
376 the hypothesis of a relaxation in selection pressure on islands due to a change in the parasite
377 community, we expect the $\Delta Pn/Ps$ to be higher on island species compared to the mainland
378 ones and, therefore, the species origin (i.e., mainland or island) to be significant. In this model,
379 we used the Phylogenetic Generalized Least Squares model (PGLS; implemented in the “nlme”
380 packages; Pinheiro et al., 2017). This model assumes that the covariance between species
381 follows a Brownian motion evolution process along the phylogeny (implemented using the
382 “corBrownian” function from the “ape” package; Paradis and Schliep, 2019). The species
383 phylogeny was estimated using mitochondrial genes and a maximum likelihood inference
384 implemented in IQTREE (model GTR+Gamma and ultrafast bootstrap; Nguyen et al., 2014;
385 median of 11,134 bp analysed per species). The phylogeny with the bootstrap support is
386 provided as supplementary material.

387 All the statistical analyses were performed using R (R Core Team, 2018), and dplyr package
388 (Wickham, 2016). Graphical representations were done using ggplot2, ggrepel, ggpubr and
389 ggpmisc (Aphalo, 2020; Kassambara, 2018; Slowikowski et al., 2018; Wickham, 2016).

390

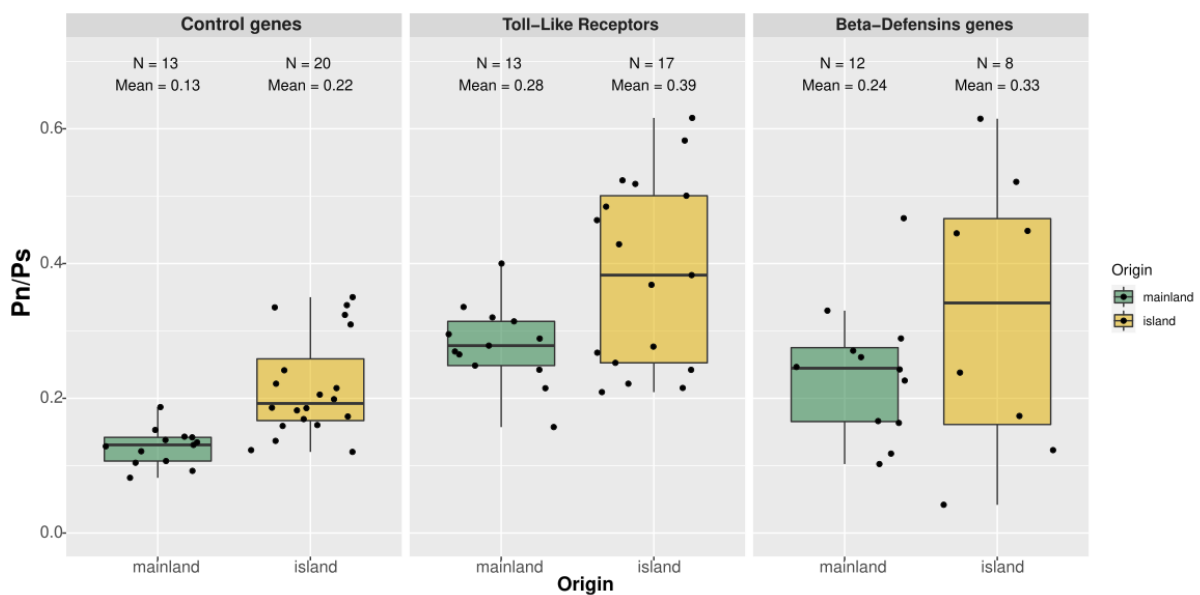
391 Results

392 For the 150 individuals (10 species with 15 individuals each) for which we generated new data
393 by targeted capture sequencing, an average of 3.3 million paired-ends reads per individual was
394 generated (Table S1). After mapping, genotyping and cleaning, we analysed 86 control and 16
395 immune genes on average per species, out of the 141 targeted genes (120 control and 21
396 immune related genes; Table S4). For the species with whole-genome sequences, we analysed
397 106 control and 20 immune genes on average per species, out of the 141 targeted genes, and
398 875 and 688 genes on average in the Database-group and Sma3s-group respectively (Table S4).

399 For the species for which full genome sequences were available, the Ps and Pn/Ps estimated
400 using the control genes reflect the Ps and Pn/Ps of the whole transcriptome (Figure S6).

401 Population genetics of BD and TLR Immune genes

402 In order to characterize the selection regimes shaping the BD and TLR polymorphisms, we
403 first analyze the variation of Pn/Ps ratios among gene categories using a linear mixed model.



404

405 Figure 3: Pn/Ps according to species origin (mainland in green and insular in orange) for
406 different gene categories under purifying selection. The number of species (N), and the mean
407 Pn/Ps are shown for each modality.

408 Model selection based on AICc as well as model selection approach based on simplification
409 with ANOVA identified the model including the origin (i.e., mainland or island) and gene
410 category without interaction (Table 2). In this model, island origin of species is associated with
411 a greater Pn/Ps (0.14 vs. 0.10; Table 3; $p < 0.01$). Gene categories corresponding to TLRs and
412 BDs showed a significantly higher Pn/Ps than control genes (Table 3; $p < 0.001$). Our statistical
413 analysis confirms that island birds have a higher Pn/Ps ratio than mainland relatives, in
414 agreement with the nearly-neutral theory of evolution. It also reveals that immune genes have
415 a higher Pn/Ps than randomly selected control genes suggesting that BD and TLR evolve under
416 a different selection regime than non-immune related genes.

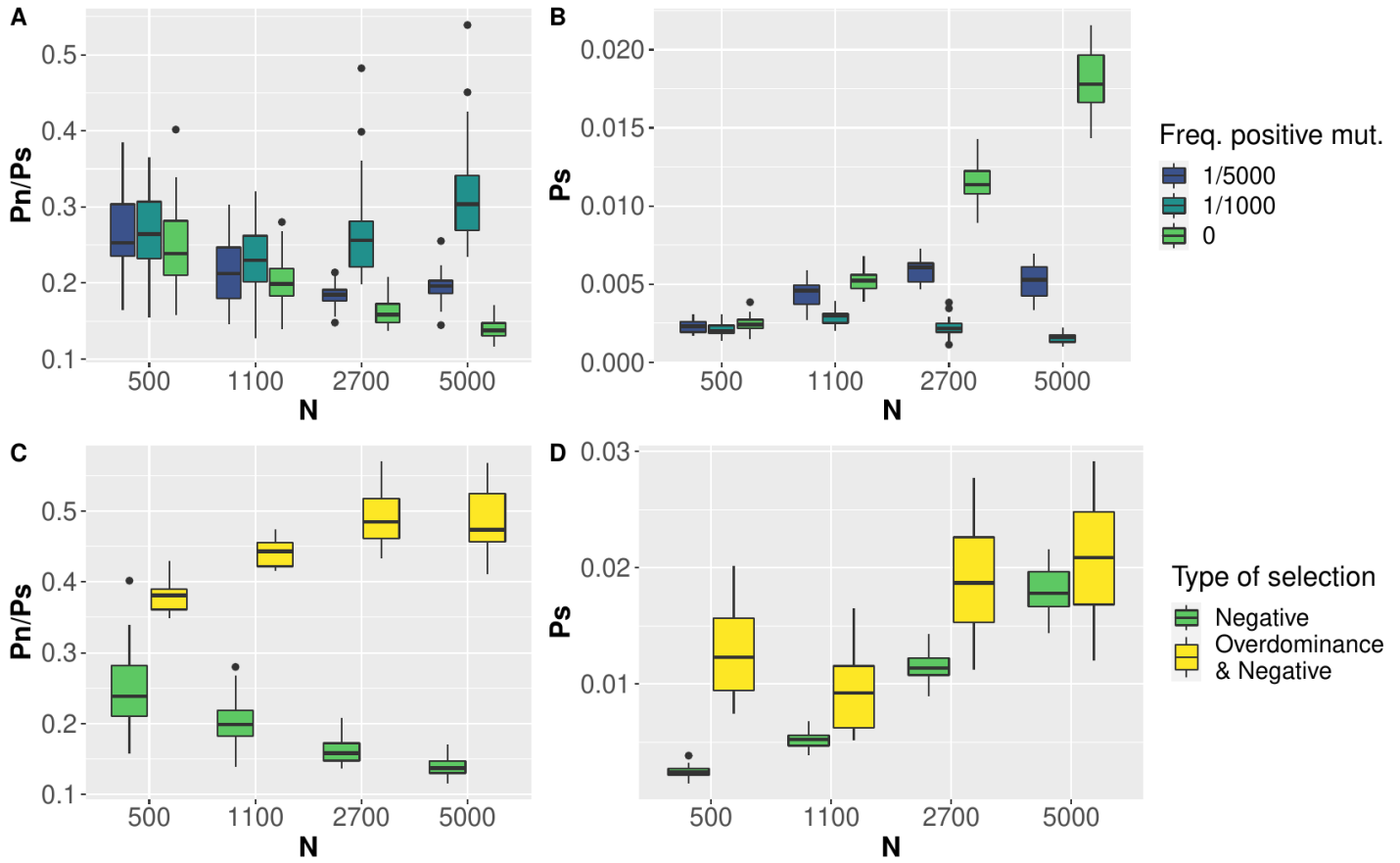
417 Next, we investigate the cause of the higher Pn/Ps of immune genes by testing three hypotheses.
418 First, we exclude a bias due to a lower number of genes in immune genes, and therefore higher
419 variance in the estimation of Pn/Ps, in immune genes. Immune genes still have significantly
420 higher Pn/Ps compared to a random subsample of control genes of comparable size (Figure S8
421 & S9). Second, the Pn/Ps of immune genes could be inflated by positive selection. It is well
422 known that immune genes are subject to frequent adaptation due to harm race evolution with
423 pathogens (Enard et al., 2016; Shultz and Sackton, 2019; Velová et al., 2018). We evaluate the
424 effect of positively selected genes on the Pn/Ps using SLiM simulations with both positively
425 and negatively selected mutations. The presence of recurrent positive selection could increase
426 the Pn/Ps leading to a higher Pn/Ps in immune genes if this category is more prone to adaptive
427 evolution (Figure 4A). However, positive selection always leads to a drastic decrease in Ps due
428 to genetic sweep effect at linked sites (Figure 4B). BDs and TLRs have a slightly higher or
429 similar Ps than control genes (Figure S9, mean Ps = 0.007, 0.004 and 0.003 for BDs, TLRs and
430 control genes respectively, effect of gene category $p < 0.1$) and, as a consequence, even if
431 positive selection is likely to have impacted the evolution of immune genes, it is not the cause
432 of the higher Pn/Ps observed here. Third, balancing selection could be present, at least
433 temporarily, in the evolution of BDs and TLRs genes (Kloch et al., 2018; Levy et al., 2020).
434 Simulation analyses confirm that balancing selection causes an increase of Ps and Pn/Ps (Figure
435 4C & 4D). However, a change in effective population size has an opposite effect on the Pn/Ps
436 according to the type of selection. In the presence of slightly deleterious mutations, Pn/Ps
437 decreases with N_e whereas it increases in the presence of balancing selection. Island birds have

438 higher Pn/Ps ratios than mainland birds for BDs and TLRs. Therefore, we can rule out
439 balancing selection as the main factor explaining the high Pn/Ps of immune genes because, in
440 this case, Pn/Ps of island birds should be lower. The last possible explanation we can think of
441 is a relaxed selection of immune genes. It is likely that immune genes are overall less
442 constrained than the control genes. It has been shown that evolutionary constraints are more
443 related to gene expression than to function (Drummond et al., 2005; Drummond and Wilke,
444 2008) and therefore, functionally important genes could still have a high Pn/Ps.

445 Overall, our analyses do not support a strong impact of ongoing adaptive mutation or balancing
446 selection on BDs and TLRs. However, these immune genes do not evolve as the random genes
447 not involved in immune functions and present a significantly higher Pn/Ps of 0.20 ($p < 0.001$;
448 Table 3).

449 *No evidence of a reduced impact of the parasite communities on the polymorphism pattern*
450 *immunes genes in island birds*

451 For BDs and TLRs, the best model selected includes the origin (i.e., mainland or island) and
452 gene category without interaction (see above and Table 2). This model has no interaction
453 between origin and gene categories invalidating the hypothesis of a reduced parasite
454 communities on island (Figure 2).



456 [Figure 4](#): Neutral polymorphism (P_s) and ratio of selected over neutral polymorphism (P_n/P_s)
 457 estimated from SLiM simulations. A) P_n/P_s as a function of population size, N and B) P_s as a
 458 function of N . In both A and B, colour indicates the frequency of positively selected mutation
 459 compare to deleterious mutation. C) P_n/P_s as a function of N and D) P_s as a function of N . In
 460 both C and D, yellow indicates simulations with overdominance mutation ($h = 1.2$) and
 461 negatively selected mutations and green indicates simulations with only negatively selected
 462 mutations.

463

464 [Table 2](#): Statistical model explaining P_n/P_s variation of Toll-Like Receptors, Beta-Defensins
 465 genes, and control genes. The p-values of ANOVA test between simpler models are not
 466 reported if a more complex model explains a larger proportion of the variance.

467

Model		Model selection by AIC			ANOVA test			
n°	Details	AICc	$\Delta AICc$	Likelihood	n° 1	2	3	4
1	$P_n/P_s \sim 1 + \text{category} + \text{origin} +$	-5.39	8.83	0.01		0.63		

	category *origin					
2	Pn/Ps~ 1+ category +origin	-14.22	0	1	0.002	3.71E-05
3	Pn/Ps~1+ category	-11.8	2.42	0.3		
4	Pn/Ps~1+ origin	-6.83	7.39	0.02		
5	Pn/Ps~1	-6.44	7.78	0.02		

468

469 **Table 3:** Summary of the best statistical model selected using AICc explaining variation in
470 Pn/Ps in control genes, Toll-Like receptors and Beta-Defensins genes under purifying selection
471 with origin, gene category parameters. * indicates significances : * < 0.05; ** < 0.01; *** <
472 0.001.

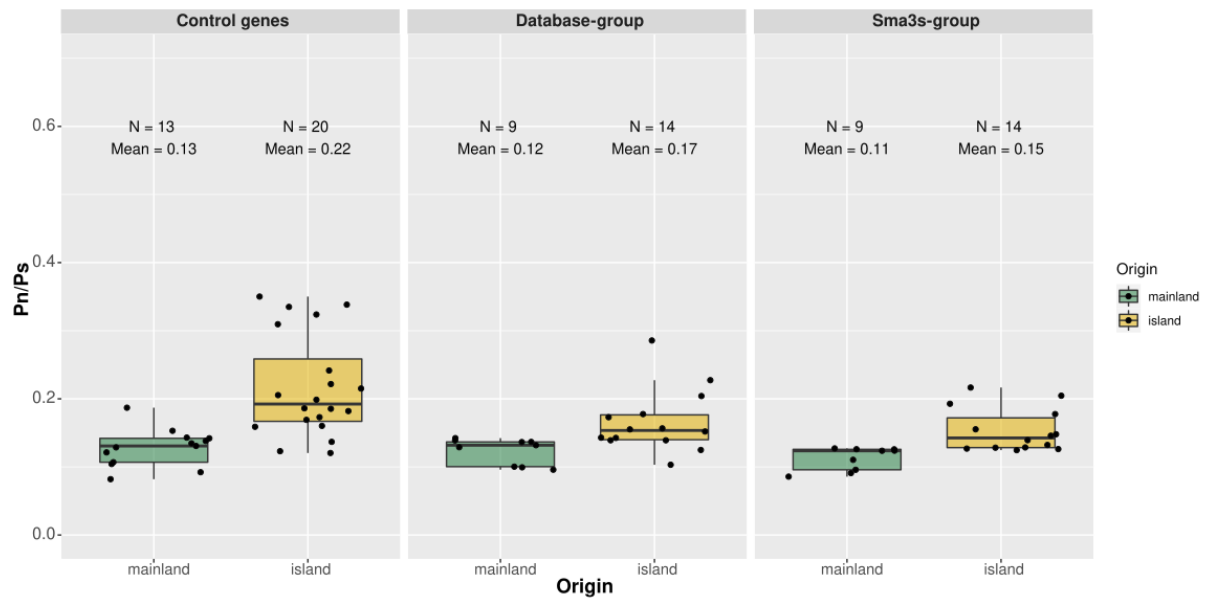
Model	Parameters		Estimate	P.value	
	Origin	Category			
<i>Origin</i> <i>and Gene</i> <i>category</i> <i>(n°2)</i>	<i>Intercept mainland</i>	<i>Control genes</i>	<i>0.10</i>	<i>2.65E-02</i>	<i>*</i>
	<i>island</i>		<i>0.14</i>	<i>4.56E-03</i>	<i>**</i>
		<i>Toll-Like Receptors</i>	<i>0.20</i>	<i>7.43E-05</i>	<i>***</i>
		<i>Beta-Defensins genes</i>	<i>0.20</i>	<i>3.16E-04</i>	<i>***</i>

473

474 For larger sets of genes, identified using an automatic pipeline and gene annotation, model
475 selection based on AICc and simplification with ANOVA (Table S5, S8) identified models
476 that included origine parameters which associated a higher Pn/Ps of at least 0.07 for island
477 species ($p < 0.001$; Table S6, S7, S9, S10, Figure 5). Selection model by simplification with
478 ANOVA identified models with interaction effect between origin and gene category associated
479 with a reduced Pn/Ps for immune genes of island species that invalidate our hypothesis (Table
480 S7, S10).

481

482 The alternative statistical approach using the difference between Pn/Ps of immune genes and
483 control genes (Δ Pn/Ps) as dependent variable, and species origin as explanatory variable under
484 a PGLS framework lead to similar results. Island was never associated to a statistically higher
485 Δ Pn/Ps (table S2) providing no support for an increased relaxed selection of immune genes in
486 island species.

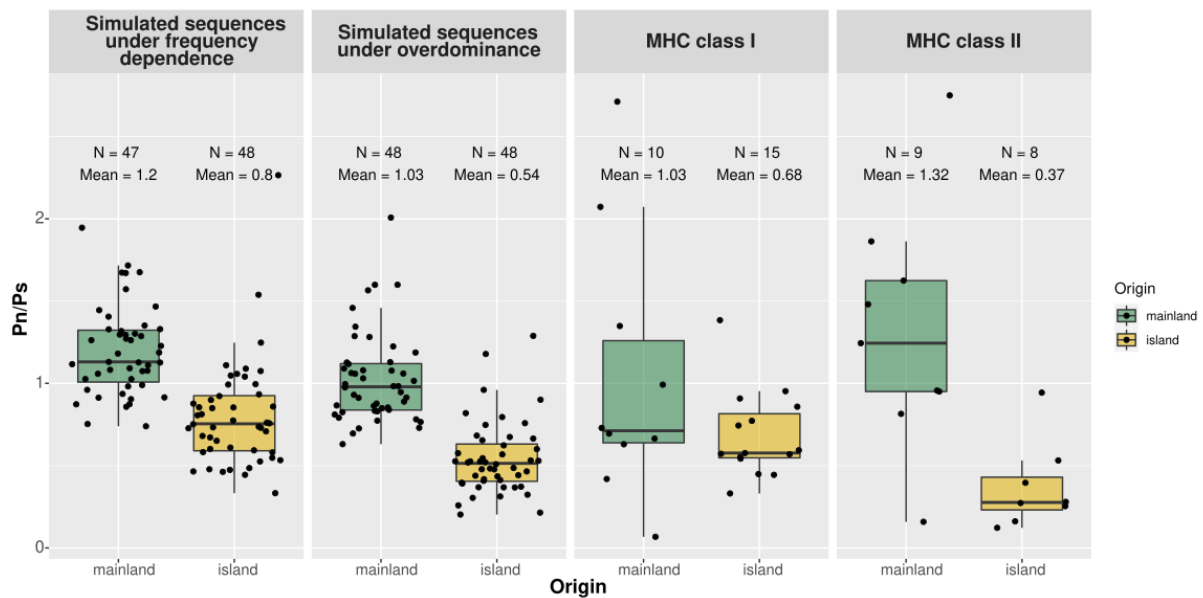


487

488 **Figure 5:** Boxplot of Pn/Ps according to species origin (mainland in green and insular in orange)
 489 for different gene categories under purifying selection. The number of individuals (N), and the
 490 mean Pn/Ps are shown for each modality.

491 *Genes under balancing selection*

492 First, we estimated the effect of population size variation on the Pn/Ps of the genes evolving
 493 under balancing selection by simulating sequences under frequency dependent or
 494 overdominance selection using SLiM (see Methods and Supplementary Methods). The
 495 simulation under frequency dependence selection revealed an average Pn/Ps equal to **0.8** for
 496 island species and **1.2** for mainland species (Figure 6). Under overdominance, simulated
 497 sequences from island and mainland populations respectively have an average Pn/Ps equal to
 498 **0.54** and **1.03** (Figure 6).



499

500 Figure 6: Boxplot of Pn/Ps according to species origin (mainland in green and insular in orange)
 501 for different gene categories under balancing selection. The number of species (N), and the
 502 mean Pn/Ps are shown for each modality. The control groups correspond to the results obtained
 503 from simulated sequence via SLiM (see Methods and Supplementary Methods Simulation of
 504 control genes under balancing selection).

505 Using simulations under frequency dependence selection as well as simulations under the
 506 overdominance, model selection by AIC identifies the model with origin as the best, contrary
 507 to the method by simplification with ANOVA which identified the full model therefore
 508 including significant interaction between origin and genes category (Table 4). This interaction
 509 effect is significant for the MHC II ($p < 0.05$, Table S12) but not for MHC I. As expected,
 510 island species have a significantly lower Pn/Ps in MHC genes compared to mainland species
 511 ($p < 0.01$; except for the full model based on control genes evolving under overdominance
 512 Table S12).

513 Table 4: Statistical model explaining Pn/Ps variation of genes under balancing selection (i.e
 514 MHC class I and II), and simulated sequences under i) frequency dependence or ii)
 515 overdominance. The p-values of ANOVA test between simpler models are not reported if a
 516 more complex model explains a larger proportion of the variance.

517

Model		Model selection by AIC			ANOVA test				
Type of balancing selection	n°	Details	AICc	$\Delta AICc$	Likelihood	n°1	2	3	4

Frequency dependence	1	Pn/Ps~1+ category +origin+ category *origin	157.17	5.62	0.06	0.019
	2	Pn/Ps~1+ category +origin	157.85	6.31	0.04	
	3	Pn/Ps~1+ category	187.58	36.04	0.00	
	4	Pn/Ps~1+ origin	151.54	0.00	1.00	
	5	Pn/Ps~1	180.52	28.97	0.00	
Overdominance	1	Pn/Ps~1+ category +origin+ category *origin	140,56	8,50	0,01	0.024
	2	Pn/Ps~1+ category +origin	140,56	8,50	0,01	
	3	Pn/Ps~1+ category	185,91	53,85	0,00	
	4	Pn/Ps~1+ origin	132,05	0,00	1,00	
	5	Pn/Ps~1	177,54	45,49	0,00	

518

519 **Discussion**

520 On oceanic islands, the depauperate parasite community is expected to lead to a relaxation of
521 selection on the immune system. In this study, we found support for such an effect, but only on
522 MHC class II genes and using simulated sequences under balancing selection as control. No
523 effect was detected for MHC class I genes nor for innate immune genes (TLRs and BDs),
524 evolving under purifying selection. On these gene sets, increased drift effects on island
525 populations limit the efficacy of selection in accordance with the nearly-neutral theory (Ohta,
526 1992). The ability to distinguish between the selective and nearly-neutral processes (relaxed
527 selection due to environmental change vs. drift) could only be achieved by our approach of
528 using random genes (i.e., “control genes”) to estimate the genome-wide effect of potential
529 variation in effective population size between populations.

530 *Effects of effective population size variation*

531 Our results support the nearly-neutral theory of evolution for those genes under purifying
532 selection, whereby strong genetic drift acting on small island populations reduces the efficacy
533 of natural selection, leading to an increase in non-synonymous nucleotide diversity compared
534 to the mostly neutral, synonymous nucleotide diversity (i.e., Pn/Ps; Ohta, 1992). This is
535 materialized by a genome-wide increase in frequency of weakly deleterious mutations
536 (Kutschera et al., 2020; Leroy et al., 2021b; Loire et al., 2013; Robinson et al., 2016; Rogers
537 and Slatkin, 2017).

538 For genes evolving under balancing selection, we performed simulations under the hypotheses
539 of overdominance (heterozygote advantage) or frequency dependence (rare-allele advantage).
540 Our results showed reduced Pn/Ps for smaller population sizes (Figure 6, S10, S11). This
541 simulation confirmed our expectations (Figure 2) that a reduction in the efficacy of selection
542 results in a decrease in the frequency of non-synonymous polymorphism, as, under normal
543 circumstances, selection maintains those mutations at intermediate frequencies. It also matches
544 what we obtained in the empirical results, where both MHC classes I and II had a reduced Pn/Ps
545 in island birds. This result supports that the fitness effect of having non-synonymous
546 polymorphisms segregating at high frequencies is not strong enough to counteract entirely the
547 effect of genetic drift on islands, therefore extending the nearly-neutral theory to the
548 overdominance type of selection.

549

550 *Effects of selection on immune genes*

551 For immune genes, we try to characterize the nature of the selection acting on BDs and TLRs
552 genes. Comparing those genes with control genes and using simulations, we were able to rule
553 out that directional positive selection and balancing selection had a major impact shaping the
554 polymorphism of these immune genes. In contrast, the pattern of Pn/Ps between island and
555 mainland populations is in line with the effect of purifying selection in the presence of slightly
556 deleterious mutation. However, no effect was detected on insular species, beyond what could
557 be attributed to genetic drift. This is in line with the result of Gonzalez-Quevedo et al. (2015b)
558 and Grueber et al. (2013) who found that TLR genetic diversity was mostly influenced by
559 genetic drift. At first sight, this result seems not in line with the fact that island parasite
560 communities are less diverse (Beadell et al., 2006; Loiseau et al., 2017; Maria et al., 2009;

561 Pérez-Rodríguez et al., 2013; but see Illera et al., 2015). However, a reduced pathogens number
562 has also been found to be associated with a higher prevalence in birds and reptiles from the
563 Macaronesian archipelago (Illera and Perera, 2020). Therefore, these two patterns, i.e. a less
564 diverse pathogen's community on islands with a higher prevalence, could still imply a strong
565 selection pressure on immune genes.

566 In contrast, for [MHC genes that unambiguously evolve under balancing selection](#), MHC class
567 II genes presented a reduction in non-synonymous polymorphism larger than the effects of drift
568 alone, when simulated sequences are used as control. This was the only case where a role for
569 relaxed selection pressures in the molecular evolution of immune genes could be invoked.

570 Our results are in accordance with the hypothesis of Lee (2006), which proposes that innate
571 and acquired immunity may exhibit distinct responses to changes in pressures due to different
572 costs and benefits. However, our result contrasts with the study of Santonastaso et al. (2017)
573 that identified no change in selection pressures on MHC II genes in a lizard species, concluding
574 that their evolution was mostly governed by drift. Similarly, Agudo et al. (2011) also found a
575 prominent role for genetic drift over selection in the evolution of MHC II genes in the Egyptian
576 vulture (*Neophron percnopterus*).

577 Our results rely on simulations that may be affected by the choice of the parameter values.
578 First, we performed simulations using a fixed effective population size (N_e) estimated from the
579 polymorphism data. Using others values of N_e had a weak impact on the relative difference
580 between island and mainland species for the overdominance type of selection (Figure S10,
581 S11). Secondly, we simulated two types of selection, namely overdominance (Doherty and
582 Zinkernagel, 1975) and frequency dependence (Slade and McCallum, 1992), but it has been
583 argued that the maintenance of MHC polymorphism could be the result of fluctuating selection
584 (Hill, 1991). Additionally, recombination and gene conversion has also been put forward as a
585 mechanism responsible for generating diversity (Spurgin et al., 2011). Therefore, our results
586 for the MHC II, which is based on the relative difference between P_n/P_s of island and mainland
587 species comparing empirical and simulated data, should be taken cautiously as their
588 significance can be dependent on the specific parameters that we used, although we did our
589 best to select a realistic range of parameters.

590 The observed difference between MHC class I and II could be explained by their different
591 pathogen targets: MHC class I genes are primarily involved in the recognition of intracellular

592 pathogens (Kappes and Strominger, 1988), while MHC class II genes are directly involved in
593 the recognition of extracellular pathogens (Bjorkman and Parham, 1990). These differences
594 could lead to variable selection pressures depending on the extracellular versus intracellular
595 parasite communities present on islands. In addition, the relaxed selection pressures on MHC
596 II genes from adaptive immunity is in line with a reduction in acquired immunity parameters
597 observed by Lobato et al. (2017) that used partly the same sets of species.

598 As a perspective of our work, we should mention that there is an extensive variation in the
599 number of MHC gene copies across the avian phylogeny (Minias et al., 2019; O'Connor et al.,
600 2020). Particularly, it was recently discovered that Passerines have a very dynamic evolution
601 of duplication/loss events compared to other birds (Minias et al., 2019). Here, we used the two
602 copies of MHC gene I and II currently annotated in the collared flycatcher genome as target
603 sequences for our targeted-capture sequencing. The future improvement of genome assembly,
604 thanks to the development of long-reads technology (Peona et al., 2021, 2018), will certainly
605 help to precisely annotate all MHC copies and to study the whole repertoire of MHC genes.

606 *Consequences of drift effect and selection on immunity*

607 The potential relaxation of the natural selection acting on immune genes in island species is
608 expected to reduce immune functions and increase susceptibility of island populations to
609 pathogens. This is true even if this relaxation is only the consequence of a reduction in the
610 effective population size and not caused by a reduction of the pressure exerted by the parasitic
611 community. This is in line with the results of Hawley et al. (2005) and Belasen et al. (2019)
612 who showed that a decrease in diversity of immune loci (MHC II or through immune proxy)
613 was associated with a reduction in immune functions. It should be noted that even if migration
614 rate is reduced on islands, sedentary and endemic island species are not completely free from
615 the exposure of exogen pathogens through migratory birds (Levin et al., 2013).

616 As a final remark, we would like to stress that more research is still needed (i) to ascertain both
617 selection pressures on innate and adaptive immune responses and the load of deleterious
618 mutations due to drift, also identified by an increasing body of work (Loire et al., 2013;
619 Robinson et al., 2016; Rogers and Slatkin, 2017; Kutschera et al., 2020; Leroy et al., 2021b),
620 and (ii) to describe island parasite communities. To date, most of the studies investigated
621 intracellular parasite communities on islands, and more specifically haemosporidian parasites,
622 avian pox and coccidian parasites (Cornuault et al., 2012; Illera et al., 2015, 2008; Ishtiaq et

623 al., 2010; Loiseau et al., 2017; Martinez et al., 2015; Padilla et al., 2017; Pérez-Rodríguez et
624 al., 2013; Silva-Iturriza et al., 2012), whereas very few evaluated the extracellular parasite
625 diversity, such as helminths (Nieberding et al., 2006) but see the review of Illera and Perera
626 (2020) for reptiles. Metabarcoding of parasites is a new technique to evaluate at the same time
627 both communities of intracellular and extracellular parasites (Bourret et al., 2021) and might
628 therefore be a promising approach to evaluate their communities in island and mainland
629 populations.

630 *Conclusion*

631 Our comparative population genomics study has investigated the combined effects of drift and
632 selection on immune genes from island and mainland passerines. The study of synonymous
633 and non-synonymous polymorphism of these genes confirmed that island species, with smaller
634 population sizes than their mainland counterparts, were more impacted by drift, which induces
635 a load of weakly deleterious mutations in their genome. Indeed most of the genes studied here
636 involved in the immune response do not show a statistically different pattern from control
637 genes. Only MHC II genes, involved in the recognition of extracellular pathogens, showed a
638 reduction in their non-synonymous polymorphism in island species. This response, which may
639 be attributed to reduced selection pressures on these genes, could be associated with the
640 suspected reduced parasitic communities on islands. The increased load of deleterious
641 mutations as well as the potential relaxed selection pressures on MHC II support the reduced
642 immune functions of island species, which could be added to the list of other convergent
643 responses of the island syndrome.

644 *Data availability*

645 Datasets, scripts, supplementary figures and texts are available on figshare :
646 <https://figshare.com/s/ab7004cc2f4415b4058f>. The reads newly generated for this study have
647 been deposited in the NCBI Sequence Read Archive under the bioproject PRJNA724656.

648 *Acknowledgments*

649 In Gabon, we thank the Director and the guides of the Lekedi Park, Marie Charpentier for her
650 help in organizing the expedition, and Alexandre Vaz for field assistance and outreach work.
651 In São Tomé and Príncipe, we thank the Directorate of the Environment and the Department
652 for Nature Conservation, its directors—Arlindo Carvalho and Victor Bonfim—Guilhermino,
653 the Association Monte Pico, its president Luis Mário, and its members. Philippe Perret, Octávio

654 Veiga, Bikegila, and Yelli provided invaluable assistance in the field. Permissions for
655 fieldwork were given by the authorities of São Tomé and Príncipe and Gabon (CENAREST
656 authorization No. AR0053/12/MENESTFPRSCJS/ CENAREST/CG/CST/CSAR). Permits for
657 the Canary Islands were provided by the Regional Government (Ref.: 2012/0710), and the
658 Cabildo of La Palma and Tenerife. In Montpellier, we thank the blue tit team
659 (<https://oreme.org/observation/ecopop/mesanges/>) for the capture of the individuals used in
660 this study. The analyses benefited from the Montpellier Bioinformatics Biodiversity (MBB)
661 platform services. This research was conducted in the scope of the international twin-lab “LIA
662 – Biodiversity and Evolution” between CIBIO (Portugal) and ISEM and CEFE-CNRS
663 (France).

664 *Funding information*

665 This research was funded by the Labex CeMEB (project ISLAND IMMUNITY) for BN, the
666 ANR (BirdIslandGenomic project, ANR-14-CE02-0002) for MB, BN, CL and CD, the
667 National Geographic Society (Grant/Award Number:W251-12), the British Ecological Society
668 (Grant/Award Number: 369/4558), the portuguese Foundation for Science and Technology
669 under the PTDC/BIA-EVL/29390/2017 “DEEP” Research Project for MM, RC and CL and
670 the Spanish Ministry of Science, Innovation and Universities, and the European Regional
671 Development Fund (Ref.: PGC2018-097575-B-I00) for JCI.

672 **References**

- 673 Agudo, R., Alcaide, M., Rico, C., Lemus, J.A., Blanco, G., Hiraldo, F., Donázar, J.A., 2011.
674 Major histocompatibility complex variation in insular populations of the Egyptian
675 vulture: inferences about the roles of genetic drift and selection. *Mol. Ecol.* 20, 2329–
676 2340. <https://doi.org/10.1111/j.1365-294X.2011.05107.x>
- 677 Akira, S., 2003. Toll-like receptor signaling. *J Biol Chem* 278, 38105–38108.
- 678 Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. Innate immunity.
679 *Mol. Biol. Cell.*
- 680 Aphalo, P.J., 2020. ggpmisc: Miscellaneous Extensions to “ggplot2”(R package version 0.3.
681 6.
- 682 Baeckens, S., Van Damme, R., 2020. The island syndrome. *Curr. Biol.* 30, R338–R339.
- 683 Bates, D.M., Maechler, M., Bolker, B., Walker, S., 2012. Package ‘lme4.’ CRAN R Found
684 *Stat Comput.*
- 685 Beadell, J.S., Atkins, C., Cashion, E., Jonker, M., Fleischer, R.C., 2007. Immunological
686 change in a parasite-impoverished environment: divergent signals from four island
687 taxa. *PLoS One* 2:e896.
- 688 Beadell, J.S., Ishtiaq, F., Covas, R., Melo, M., Warren, B.H., Atkinson, C.T., Bensch, S.,
689 Graves, G.R., Jhala, Y.V., Peirce, M.A., 2006. Global phylogeographic limits of
690 Hawaii’s avian malaria. *Proc R Soc Lond B Biol Sci* 273, 2935–2944.
- 691 Belasen, A.M., Bletz, M.C., S, L.D., Toledo, L.F., James, T.Y., 2019. Long-term habitat

692 fragmentation is associated with reduced MHC IIB diversity and increased infections
693 in amphibian hosts. *Front Ecol Evol* 6.

694 Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: what have we
695 learned about natural selection in 15 years? *J Evol Biol* 16, 363–377.

696 Bjorkman, P.J., Parham, P., 1990. Structure, function, and diversity of class I major
697 histocompatibility complex molecules. *Annu Rev Biochem* 59, 253–288.

698 Blondel, J., 2000. Evolution and ecology of birds on islands: trends and prospects. *Vie*
699 *MilieuLife Environ.* 205–220.

700 Bourret, V., Gutiérrez López, R., Melo, M., Loiseau, C., 2021. Metabarcoding options to
701 study eukaryotic endoparasites of birds. *Ecol. Evol.* 11, 10821–10833.

702 Boyce, M.S., 1984. Restitution of gamma-and k-selection as a model of density-dependent
703 natural selection. *Annu Rev Ecol Syst* 15, 427–447.

704 Breuer, K., Foroushani, A.K., Laird, M.R., Chen, C., Sribnaia, A., Lo, R., Winsor, G.L.,
705 Hancock, R.E., Brinkman, F.S., Lynn, D.J., 2013. InnateDB: systems biology of
706 innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids*
707 *Res* 41:D1228–D1233.

708 Buffalo, V., 2021. Quantifying the relationship between genetic diversity and population size
709 suggests natural selection cannot explain Lewontin’s paradox. *eLife* 10, e67509.
710 <https://doi.org/10.7554/eLife.67509>

711 Castellano, D., James, J., Eyre-Walker, A., 2018. Nearly neutral evolution across the
712 *Drosophila melanogaster* genome. *Mol. Biol. Evol.* 35, 2685–2694.

713 Chapman, H., JR, O, H., AS, K., RH, C., RL, W., J., 2016. The evolution of innate immune
714 genes: purifying and balancing selection on β -defensins in waterfowl. *Mol Biol Evol*
715 33, 3075–3087.

716 Charlesworth, J., Eyre-Walker, A., 2008. The McDonald–Kreitman test and slightly
717 deleterious mutations. *Mol Biol Evol* 25, 1007–1015.

718 Chen, J., Glémin, S., Lascoux, M., 2020. From drift to draft: how much do beneficial
719 mutations actually contribute to predictions of Ohta’s slightly deleterious model of
720 molecular evolution? *Genetics* 214, 1005–1018.

721 Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ
722 preprocessor. *Bioinformatics* 34:i884–i890.

723 Corcoran, P., Gossman, T.I., Barton, H.J., Slate, J., Zeng, K., 2017. Determinants of the
724 Efficacy of Natural Selection on Coding and Noncoding Variability in Two Passerine
725 Species. *Genome Biol. Evol.* 9, 2987–3007. <https://doi.org/10.1093/gbe/evx213>

726 Cornuault, J., Bataillard, A., Warren, B.H., Lootvoet, A., Mirleau, P., Duval, T., Milá, B.,
727 Thébaud, C., Heeb, P., 2012. The role of immigration and in-situ radiation in
728 explaining blood parasite assemblages in an island bird clade. *Mol. Ecol.* 21, 1438–
729 1452.

730 Covas, R., 2012. Evolution of reproductive life histories in island birds worldwide. *Proc R*
731 *Soc B Biol Sci* 279, 1531–1537.

732 Doherty, P.C., Zinkernagel, R.M., 1975. Enhanced immunological surveillance in mice
733 heterozygous at the H-2 gene complex. *Nature* 256, 50–52.

734 Doutrelant, C., Paquet, M., Renoult, J.P., Grégoire, A., Crochet, P.-A., Covas, R., 2016.
735 Worldwide patterns of bird colouration on islands. *Ecol Lett* 19, 537–545.

736 Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., Arnold, F.H., 2005. Why highly
737 expressed proteins evolve slowly. *Proc. Natl. Acad. Sci.* 102, 14338–14343.

738 Drummond, D.A., Wilke, C.O., 2008. Mistranslation-induced protein misfolding as a
739 dominant constraint on coding-sequence evolution. *Cell* 134, 341–352.

740 Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A.,
741 Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., 2012. The genomic

742 landscape of species divergence in *Ficedula* flycatchers. *Nature* 491, 756–760.
743 Enard, D., Cai, L., Gwennap, C., Petrov, D.A., 2016. Viruses are a dominant driver of protein
744 adaptation in mammals. *elife* 5, e12469.
745 Eyre-Walker, A., Keightley, P.D., 2007. The distribution of fitness effects of new mutations.
746 *Nat. Rev. Genet.* 8, 610–618.
747 Fijarczyk, A., Dudek, K., Babik, W., 2016. Selective Landscapes in new Immune Genes
748 Inferred from Patterns of Nucleotide Variation. *Genome Biol Evol* 8, 3417–3432.
749 Frankham, R., 1997. Do island populations have less genetic variation than mainland
750 populations? *Heredity* 78, 311–327.
751 Fu, L., Niu, B., Zhu, Z., Wu, S., Li, W., 2012. CD-HIT: accelerated for clustering the next-
752 generation sequencing data. *Bioinformatics* 28, 3150–3152.
753 Garamszegi, L.Z., 2006. The evolution of virulence and host specialization in malaria
754 parasites of primates. *Ecol Lett* 9, 933–940.
755 Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing.
756 Gonzalez-Quevedo, C., Phillips, K.P., Spurgin, L.G., Richardson, D.S., 2015a. 454 screening
757 of individual MHC variation in an endemic island passerine. *Immunogenetics* 67,
758 149–162. <https://doi.org/10.1007/s00251-014-0822-1>
759 Gonzalez-Quevedo, C., Spurgin, L.G., Illera, J.C., Richardson, D.S., 2015b. Drift, not
760 selection, shapes toll-like receptor variation among oceanic island populations. *Mol.*
761 *Ecol.* 24, 5852–5863.
762 Grant, P.R., 1965. The adaptive significance of some size trends in island birds. 355–367,
763 *Evolution*.
764 Grueber, C.E., Wallis, G.P., Jamieson, I.G., 2014. Episodic positive selection in the evolution
765 of avian toll-like receptor innate immunity genes. *PloS One* 9, e89632.
766 Grueber, C.E., Wallis, G.P., Jamieson, I.G., 2013. Genetic drift outweighs natural selection at
767 toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened
768 species. *Mol. Ecol.* 22, 4470–4482.
769 Guéguen, L., Gaillard, S., Boussau, B., Gouy, M., Groussin, M., Rochette, N.C., Bigot, T.,
770 Fournier, D., Pouyet, F., Cahais, V., Bernard, A., Scornavacca, C., Nabholz, B.,
771 Haudry, A., Dachary, L., Galtier, N., Belkhir, K., Dutheil, J.Y., 2013. Bio++:
772 Efficient Extensible Libraries and Tools for Computational Molecular Evolution.
773 *Mol. Biol. Evol.* 30, 1745–1750. <https://doi.org/10.1093/molbev/mst097>
774 Hale, K.A., Briskie, J.V., 2007. Decreased immunocompetence in a severely bottlenecked
775 population of an endemic New Zealand bird. *Anim. Conserv.* 10, 2–10.
776 Haller, B.C., Messer, P.W., 2017. SLiM 2: Flexible, interactive forward genetic simulations.
777 *Mol Biol Evol* 34, 230–240.
778 Hawley, D.M., Sydenstricker, K.V., Kollias, G.V., Dhondt, A.A., 2005. Genetic diversity
779 predicts pathogen resistance and cell-mediated immunocompetence in house finches.
780 *Biol Lett* 1, 326–329.
781 Hill, A.V., 1991. HLA associations with malaria in Africa: some implications for MHC
782 evolution, in: *Molecular Evolution of the Major Histocompatibility Complex*.
783 Springer, pp. 403–420.
784 Hochberg, M.E., Møller, A.P., 2001. Insularity and adaptation in coupled victim–enemy
785 associations. *J Evol Biol* 14, 539–551.
786 Illera, J.C., Emerson, B.C., Richardson, D.S., 2008. Genetic characterization, distribution and
787 prevalence of avian pox and avian malaria in the Berthelot’s pipit (*Anthus berthelotii*)
788 in Macaronesia. *Parasitol Res* 103, 1435–1443.
789 Illera, J.C., Fernández-Álvarez, Á., Hernández-Flores, C.N., Foronda, P., 2015. Unforeseen
790 biogeographical patterns in a multiple parasite system in Macaronesia. *J. Biogeogr.*
791 42, 1858–1870.

792 Illera, J.C., Perera, A., 2020. Where are we in the host-parasite relationships of native land
793 vertebrates in Macaronesia? *Ecosistemas*.
794 Institute, B., 2019. “Picard Toolkit”, Broad institute, GitHub repository. Picard Toolkit.
795 Ishtiaq, F., Clegg, S.M., Phillimore, A.B., Black, R.A., Owens, I.P., Sheldon, B.C., 2010.
796 Biogeographical patterns of blood parasite lineage diversity in avian hosts from
797 southern Melanesian islands. *J. Biogeogr.* 37, 120–132.
798 Jombart, T., Ahmed, I., 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP
799 data. *Bioinformatics* 27, 3070–3071.
800 Kappes, D., Strominger, J.L., 1988. Human class II major histocompatibility complex genes
801 and proteins. *Annu Rev Biochem* 57, 991–1028.
802 Kassambara, A., 2018. ggpubr:“ggplot2” based publication ready plots. R Package Version
803 01, 7.
804 Kimura, M., 1962. On the Probability of Fixation of Mutant Genes in a Population. *Genetics*
805 47, 713–719.
806 Klein, J., 1986. Natural history of the major histocompatibility complex. Wiley.
807 Kloch, A., Wenzel, M.A., Laetsch, D.R., Michalski, O., Bajer, A., Behnke, J.M., Welc-
808 Fałęciak, R., Piertney, S.B., 2018. Signatures of balancing selection in toll-like
809 receptor (TLRs) genes—novel insights from a free-living rodent. *Sci. Rep.* 8, 1–10.
810 Kutschera, V.E., Poelstra, J.W., Botero-Castro, F., Dussex, N., Gemmell, N., Hunt, G.R.,
811 Ritchie, M.G., Rutz, C., Wiberg, R.A.W., Wolf, J.B.W., 2020. Purifying Selection in
812 Corvids Is Less Efficient on Islands. *Mol. Biol. Evol.*
813 <https://doi.org/10.1093/molbev/msz233>
814 Kuznetsova, A., Brockhoff, P.B., Christensen, R.H., 2017. lmerTest package: tests in linear
815 mixed effects models. *J Stat Softw* 82, 1–26.
816 Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O.,
817 Verhoeven, K.J., De Jager, V., Megens, H.-J., Warren, W.C., Minx, P., 2016.
818 Evolutionary signals of selection on cognition from the great tit genome and
819 methylome. *Nat. Commun.* 7, 1–9.
820 Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio,
821 A., Promerová, M., Rubin, C.-J., Wang, C., Zamani, N., 2015. Evolution of Darwin’s
822 finches and their beaks revealed by genome sequencing. *Nature* 518, 371–375.
823 Lee, J.W., Beebe, K., Nangle, L.A., Jang, J., Longo-Guess, C.M., Cook, S.A., Davisson,
824 M.T., Sundberg, J.P., Schimmel, P., Ackerman, S.L., 2006. Editing-defective tRNA
825 synthetase causes protein misfolding and neurodegeneration. *Nature* 443, 50–55.
826 <https://doi.org/10.1038/nature05096>
827 Lee, K.A., 2006. Linking immune defenses and life history at the levels of the individual and
828 the species. *Integr Comp Biol* 46, 1000–1015.
829 Leroy, T., Anselmetti, Y., Tilak, M.-K., Bérard, S., Csukonyi, L., Gabrielli, M., Scornavacca,
830 C., Milá, B., Thébaud, C., Nabholz, B., 2021a. A bird’s white-eye view on avian sex
831 chromosome evolution. *Peer Community J.* 1.
832 Leroy, T., Rousselle, M., Tilak, M.-K., Caizergues, A.E., Scornavacca, C., Recuerda, M.,
833 Fuchs, J., Illera, J.C., De Swardt, D.H., Blanco, G., 2021b. Island songbirds as
834 windows into evolution in small populations. *Curr. Biol.* 31, 1303-1310. e4.
835 Levin, I.I., Zwiars, P., Deem, S.L., Geest, E.A., Higashiguchi, J.M., Iezhova, T.A., Jiménez-
836 Uzcátegui, G., Kim, D.H., Morton, J.P., Perlut, N.G., Renfrew, R.B., Sari, E.H.R.,
837 Valkiunas, G., Parker, P.G., 2013. Multiple Lineages of Avian Malaria Parasites
838 (*Plasmodium*) in the Galapagos Islands and Evidence for Arrival via Migratory Birds.
839 *Conserv. Biol.* 27, 1366–1377. <https://doi.org/10.1111/cobi.12127>
840 Levy, H., Fiddaman, S.R., Vianna, J.A., Noll, D., Clucas, G.V., Sidhu, J.K., Polito, M.J.,
841 Bost, C.A., Phillips, R.A., Crofts, S., 2020. Evidence of pathogen-induced

842 immunogenetic selection across the large geographic range of a wild seabird. *Mol.*
843 *Biol. Evol.* 37, 1708–1726.

844 Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-
845 MEM. *ArXiv Prepr. ArXiv13033997*.

846 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis,
847 G., Durbin, R., 2009. The Sequence Alignment/Map format and SAMtools.
848 *Bioinforma Oxf Engl* 25, 2078–2079.

849 Lindström, K.M., Foufopoulos, J., Pärn, H., Wikelski, M., 2004. Immunological investments
850 reflect parasite abundance in island populations of Darwin’s finches. *Proc R Soc Lond*
851 *B Biol Sci* 271, 1513–1519.

852 Lobato, E., Doutrelant, C., Melo, M., Reis, S., Covas, R., 2017. Insularity effects on bird
853 immune parameters: A comparison between island and mainland populations in West
854 Africa. *Ecol. Evol.* 7, 3645–3656.

855 Loire, E., Chiari, Y., Bernard, A., Cahais, V., Romiguier, J., Nabholz, B., Lourenço, J.M.,
856 Galtier, N., 2013. Population genomics of the endangered giant Galapagos tortoise.
857 *Genome Biol.* 14, R136. <https://doi.org/10.1186/gb-2013-14-12-r136>

858 Loiseau, C., Melo, M., Lobato, E., Beadell, J.S., Fleischer, R.C., Reis, S., Doutrelant, C.,
859 Covas, R., 2017. Insularity effects on the assemblage of the blood parasite community
860 of the birds from the Gulf of Guinea. *J. Biogeogr.* 44, 2607–2617.

861 Lomolino, M.V., 2005. Body size evolution in insular vertebrates: generality of the island
862 rule. *J. Biogeogr.* 32, 1683–1699. <https://doi.org/10.1111/j.1365-2699.2005.01314.x>

863 Losos, J.B., Ricklefs, R.E., 2009. Adaptation and diversification on islands. *Nature* 457, 830–
864 836.

865 Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S.,
866 Bensch, S., 2017. Genetic differences between willow warbler migratory phenotypes
867 are few and cluster in large haplotype blocks. *Evol Lett* 1, 155–168.

868 Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y.,
869 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo
870 assembler. *Gigascience* 1.

871 MacArthur, R.H., Wilson, E.O., 1967. The theory of island biogeography, in: *The Theory of*
872 *Island Biogeography*. Princeton university press.

873 Maria, L., Svensson, E., Ricklefs, R.E., 2009. Low diversity and high intra-island variation in
874 prevalence of avian Haemoproteus parasites on Barbados, Lesser Antilles.
875 *Parasitology* 136, 1121–1131.

876 Martinez, J., Vasquez, R.A., Venegas, C., Merino, S., 2015. Molecular characterisation of
877 haemoparasites in forest birds from Robinson Crusoe Island: is the Austral Thrush a
878 potential threat to endemic birds? *Bird Conserv. Int.* 25, 139–152.

879 Matson, K.D., 2006. Are there differences in immune function between continental and
880 insular birds? *Proc. Biol. Sci.* 273, 2267–2274.
881 <https://doi.org/10.1098/rspb.2006.3590>

882 Matson, K.D., Beadell, J.S., 2010. Infection, immunity, and island adaptation in birds.

883 Minias, P., Pikus, E., Whittingham, L.A., Dunn, P.O., 2019. Evolution of copy number at the
884 MHC varies across the avian tree of life. *Genome Biol Evol* 11, 17–28.

885 Mueller, J.C., Kuhl, H., Timmermann, B., Kempnaers, B., 2016. Characterization of the
886 genome and transcriptome of the blue tit *Cyanistes caeruleus*: polymorphisms, sex-
887 biased expression and selection signals. *Mol. Ecol. Resour.* 16, 549–561.
888 <https://doi.org/10.1111/1755-0998.12450>

889 Munoz-Mérida, A., Viguera, E., Claros, M.G., Trelles, O., Pérez-Pulido, A.J., 2014. Sma3s: a
890 three-step modular annotator for large sequence datasets. *DNA Res* 21, 341–353.

891 Nguyen, L.-T., Schmidt, H.A., Haeseler, A., Minh, B.Q., 2014. IQ-TREE: a fast and effective

892 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*
893 32, 268–274.

894 Nieberding, C., Morand, S., Libois, R., Michaux, J., 2006. Parasites and the island syndrome:
895 the colonization of the western Mediterranean islands by *Heligmosomoides polygyrus*
896 (Dujardin, 1845). *J Biogeogr* 33, 1212–1222.

897 O'Connor, E.A., Hasselquist, D., Nilsson, J.-Å., Westerdahl, H., Cornwallis, C.K., 2020.
898 Wetter climates select for higher immune gene diversity in resident, but not
899 migratory, songbirds. *Proc. R. Soc. B Biol. Sci.* 287, 20192675.
900 <https://doi.org/10.1098/rspb.2019.2675>

901 Ohta, T., 1992. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* 23,
902 263–286.

903 Ortutay, C., Vihinen, M., 2009. Identification of candidate disease genes by integrating Gene
904 Ontologies and protein-interaction networks: case study of primary
905 immunodeficiencies. *Nucleic Acids Res* 37, 622–628.

906 Padilla, D.P., Illera, J.C., Gonzalez-Quevedo, C., Villalba, M., Richardson, D.S., 2017.
907 Factors affecting the distribution of haemosporidian parasites within an oceanic
908 island. *Int. J. Parasitol.* 47, 225–235.

909 Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and
910 evolutionary analyses in R. *Bioinformatics* 35, 526–528.

911 Peona, V., Blom, M.P.K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., Haryoko, T.,
912 Jønsson, K.A., Zhou, Q., 2021. Identifying the causes and consequences of assembly
913 gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol Ecol Resour*
914 21, 263–286.

915 Peona, V., Weissensteiner, M.H., Suh, A., 2018. How complete are “complete” genome
916 assemblies?—An avian perspective. *Mol Ecol Resour* 18, 1188–1195.

917 Pérez-Rodríguez, A., Ramírez, Á., Richardson, D.S., Pérez-Tris, J., 2013. Evolution of
918 parasite island syndromes without long-term host population isolation: Parasite
919 dynamics in Macaronesian blackcaps *Sylvia atricapilla*. *Glob Ecol Biogeogr* 22,
920 1272–1281.

921 Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Willigen, B., Maintainer, R.,
922 2017. Package ‘nlme. Linear Nonlinear Mix Eff Models Version 3.

923 R Core Team, 2018. R: A language and environment for statistical computing.

924 Rando, J.C., Alcover, J.A., Illera, J.C., 2010. Disentangling Ancient Interactions: A New
925 Extinct Passerine Provides Insights on Character Displacement among Extinct and
926 Extant Island Finches. *PLOS ONE* 5:e12956.

927 Ranwez, V., Harispe, S., Delsuc, F., Douzery, E.J., 2011. MACSE: Multiple Alignment of
928 Coding SEquences accounting for frameshifts and stop codons. *PLoS One* 6:e22594.

929 Recuerda, M., Vizueta, J., Cuevas-Caballé, C., Blanco, G., Rozas, J., Milá, B., 2021.
930 Chromosome-level genome assembly of the common chaffinch (*Aves: Fringilla*
931 *coelebs*): a valuable resource for evolutionary biology. *Genome Biol. Evol.* 13,
932 evab034.

933 Robinson, J.A., Ortega-Del Vecchyo, D., Fan, Z., Kim, B.Y., Marsden, C.D., Lohmueller,
934 K.E., Wayne, R.K., 2016. Genomic flatlining in the endangered island fox. *Curr Biol*
935 26, 1183–1189.

936 Rogers, R.L., Slatkin, M., 2017. Excess of genomic defects in a woolly mammoth on
937 Wrangel island. *PLoS Genet* 13:e1006601.

938 Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for
939 multiplexed target capture. *Genome Res.* 22, 939–946.

940 Rousselle, M., Simion, P., Tilak, M.-K., Figuet, E., Nabholz, B., Galtier, N., 2020. Is
941 adaptation limited by mutation? A timescale-dependent effect of genetic diversity on

942 the adaptive substitution rate in animals. *PLoS Genet.* 16, e1008668.

943 Santonastaso, T., Lighten, J., Oosterhout, C., Jones, K.L., Foufopoulos, J., Anthony, N.M.,
944 2017. The effects of historical fragmentation on major histocompatibility complex
945 class II β and microsatellite variation in the Aegean island reptile, *Podarcis erhardii*.
946 *Ecol Evol* 7, 4568–4581.

947 She, R., Chu, J.S.-C., Uyar, B., Wang, J., Wang, K., Chen, N., 2011. genBlastG: using
948 BLAST searches to build homologous gene models. *Bioinforma Oxf Engl* 27, 2141–
949 2143.

950 Shultz, A.J., Sackton, T.B., 2019. Immune genes are hotspots of shared positive selection
951 across birds and mammals. *Elife* 8, e41815.

952 Siewert, K.M., Voight, B.F., 2020. BetaScan2: Standardized Statistics to Detect Balancing
953 Selection Utilizing Substitution Data. *Genome Biol. Evol.* 12, 3873–3877.
954 <https://doi.org/10.1093/gbe/evaa013>

955 Silva-Iturriza, A., Ketmaier, V., Tiedemann, R., 2012. Prevalence of avian haemosporidian
956 parasites and their host fidelity in the central Philippine islands. *Parasitol. Int.* 61,
957 650–657.

958 Simion, P., Belkhir, K., François, C., Veyssier, J., Rink, J.C., Manuel, M., Philippe, H.,
959 Telford, M.J., 2018. A software tool ‘CroCo’ detects pervasive cross-species
960 contamination in next generation sequencing data. *BMC Biol.* 16, 1–9.

961 Singhal, S., Leffler, E.M., Sannareddy, K., Turner, I., Venn, O., Hooper, D.M., Strand, A.I.,
962 Li, Q., Raney, B., Balakrishnan, C.N., 2015. Stable recombination hotspots in birds.
963 *Science* 350, 928–932.

964 Slade, R.W., McCallum, H.I., 1992. Overdominant vs. frequency-dependent selection at
965 MHC loci. *Genetics* 132.

966 Slowikowski, K., Schep, A., Hughes, S., Lukauskas, S., Irisson, J.-O., Kamvar, Z.N., Ryan,
967 T., Christophe, D., Hiroaki, Y., Gramme, P., 2018. Package ggrepel. Autom Position
968 Non-Overlapping Text Labels ggplot2.

969 Smeds, L., Qvarnstrom, A., Ellegren, H., 2016. Direct estimate of the rate of germline
970 mutation in a bird. *Genome Res.* gr-204669.

971 Spiess, A.-N., Spiess, M.A.-N., 2018. Package ‘qpcR’, in: *Model. Anal. Real-Time PCRdata*
972 [https://cran R-Proj. OrgwebpackagesqpcRqpcR Pdf.](https://cran.r-project.org/web/packages/qpcR/qpcR.pdf)

973 Spurgin, L.G., Van Oosterhout, C., Illera, J.C., Bridgett, S., Gharbi, K., Emerson, B.C.,
974 Richardson, D.S., 2011. Gene conversion rapidly generates major histocompatibility
975 complex diversity in recently founded bird populations. *Mol. Ecol.* 20, 5213–5225.
976 <https://doi.org/10.1111/j.1365-294X.2011.05367.x>

977 Tange, O., 2018. GNU parallel 2018.

978 van Dijk, A., Veldhuizen, E.J., Haagsman, H.P., 2008. Avian defensins. *Vet. Immunol.*
979 *Immunopathol.* 124, 1–18.

980 Van Riper III, C., Van Riper, S.G., Goff, M.L., Laird, M., 1986. The epizootiology and
981 ecological significance of malaria in Hawaiian land birds. *Ecol. Monogr.* 56, 327–
982 344.

983 Velová, H., Gutowska-Ding, M.W., Burt, D.W., Vinkler, M., 2018. Toll-like receptor
984 evolution in birds: gene duplication, pseudogenization, and diversifying selection.
985 *Mol Biol Evol* 35, 2170–2184.

986 Warren, B.H., Simberloff, D., Ricklefs, R.E., Aguilée, R., Condamine, F.L., Gravel, D.,
987 Morlon, H., Mouquet, N., Rosindell, J., Casquet, J., 2015. Islands as model systems in
988 ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecol. Lett.* 18,
989 200–217.

990 Warren, W.C., Clayton, D.F., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle,
991 S., White, S., Vilella, A.J., Fairley, S., 2010. The genome of a songbird. *Nature* 464,

992 757–762.
993 Welch, J.J., Eyre-Walker, A., Waxman, D., 2008. Divergence and Polymorphism Under the
994 Nearly Neutral Theory of Molecular Evolution. *J. Mol. Evol.* 67, 418–426.
995 <https://doi.org/10.1007/s00239-008-9146-9>
996 Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*.
997 Wikelski, M., Foufopoulos, J., Vargas, H., Snell, H., 2004. Galápagos birds and diseases:
998 invasive pathogens as threats for island species. *Ecol. Soc.* 9.
999 Wolf, J.B.W., Künstner, A., Nam, K., Jakobsson, M., Ellegren, H., 2009. Nonlinear
1000 Dynamics of Nonsynonymous (dN) and Synonymous (dS) Substitution Rates Affects
1001 Inference of Selection. *Genome Biol Evol* 1, 308–319.
1002 Zhang, G., Parker, P., Li, B., Li, H., Wang, J., 2012. The genome of Darwin’s Finch
1003 (*Geospiza fortis*). <https://doi.org/10.5524/100040>

Supplementary Text & Figures:

Evolution of immune genes in island birds: reduction in population sizes can explain island syndrome

Mathilde BARTHE, Claire DOUTRELANT, Rita COVAS, Martim MELO,

Juan Carlos ILLERA, Marie-Ka TILAK, Constance COLOMBIER, Thibault LEROY ,

Claire LOISEAU , Benoit NABHOLZ

Supplementary Methods :

Test for contamination

Analyses were performed to identify contaminations between individuals of different species. First, we generate assembly for all individuals using Megahits v.1.2.7 (Li et al. 2015). Next, we use the program CroCo (vers. 1.1; Simion et al. 2018) to identify candidates for cross-species contamination. CroCo identifies a contig from the assembly of the species X as contaminated by the species Y if the reads from the species Y map with a higher relative proportion on the contig than the reads for the species X.

Contamination tests were performed on 3765 combinations of individuals from different species. On average, 3.12% of the contigs were considered as contaminants (Figure S1). The top 20% of the combinations (i.e., a value greater than 3.09% of contaminating contigs ; Figure S1) always involve a pair of species belonging to the same genus. In this case, contamination could be difficult to identify due to the low genetic divergence between species. The remaining 80% showed low percentages of contamination (<5% of scaffolds) that may be associated with false positives (i.e., corresponding to highly conserved sequences between species). Overall, we did not detect a clear case of cross-species contamination in our dataset.

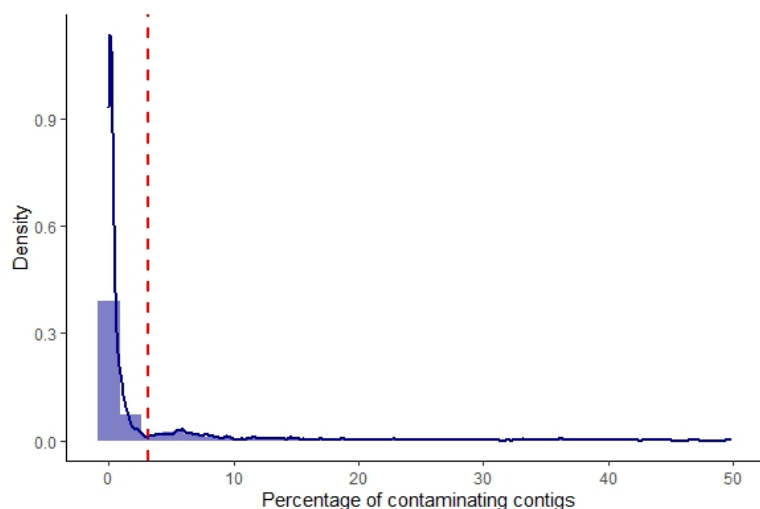


Figure S1: Distribution of the percentage of contaminating contigs. The red line represents the 80% quantile.

Homology identification

After defining the genes composition of our different groups, see Selection and identification of immune and control genes, we obtained their reference sequences from the Ensembl database for core group and control group and from the annotated genome of *F. albicollis* for Database group and Sma3s group. We also used the TLRs alignments from Velová et al. (2018).

Then, in order to identify corresponding sequences in our 11 reference genomes, we search for homologous sequences. To do so, we first translated sequences with transeq (vers. 6.6.0.0; Rice et al. 2000), then we used the program Silix (vers. 1.2.9; Miele et al. 2011) to cluster amino-acid sequences with at least 60% of similarity and a minimum overlap of 30%. For partial sequences, a minimum length of 10 amino acids or a minimum of 5% of the complete protein sequence was required.

For the immunity genes, some sequences could be assigned to several clusters due to paralogy (e.g., TLR2A and TLR2B). In the core group, we use the probabilistic approach of the profile hidden Markov models implemented in HMMER (v. 3.3.2; Mistry et al. 2013) to assign the gene to a cluster. We first created a gene-specific HMM profile from the alignments of Velová et al. (2018). Then, the candidate sequences were assigned to each homologous group based on the best score of a hmmscan search. For larger sets of genes (i.e Database-group and Sma3s-group), we remove duplicated sequences in the dataset to analyse only one.

Detection of genes under balancing selection

The Database-group and Sma3s-group most likely include immune genes subject to different selection pressures (i.e., balancing and purifying selection). To identify genes evolving under balancing selection, unfolded SNP frequency was estimated using a home-made script for *Taeniopygia guttata* using *Poephila acuticauda* as outgroup. The allelic state of *Poephila* was taken as the ancestral state of *Taeniopygia*. Then, we run the program BetaScan (vers. 2; Siewert and Voight 2020) with a sliding window of 2000pb and option '-fold -m 0.15' allowed to exclude allele with a frequency under 15% to avoid false positives to identify SNPs. We selected the top SNPs with a score above a 99.9% threshold based on the Beta* score as evolving under balancing selection (Siewert and Voight 2020). Finally, genes overlapping with the top SNPs were considered as genes evolving under balancing selection or overdominance. Respectively only 2 and 3 genes from Database-group and Sma3s-group sets were identified and removed from the analysis.

Simulation of control genes under balancing selection

Simulations were performed using SLiM software (vers. 3.3.2; Haller and Messer 2017). Sequences of 300kb with a mutation of 4.6×10^{-9} substitutions/site/generation were simulated (Smeds et al. 2016). Recombination was set to be equal to mutation rate. Three types of mutations were possible: i) neutral synonymous mutations, ii) non-synonymous mutations

with a Distribution of Fitness Effect (DFE) following a gamma law of mean = -0.025 and shape = 0.3, which corresponds to the DFE estimated in Passerines by Rousselle et al. (2020), iii) non-synonymous mutations under balancing selection with an effect on fitness initially set at 0.01 but re-estimated by the program at each generation according to the mutation frequency in the population, thus including a frequency-dependent effect.

We simulate a coding sequence organization where positions one and two of the codons were considered as non-degenerated sites with the two non-synonymous types of mutations previously described are possible in proportion of 1 mutation under balancing selection for 6 randomly sampled in the DFE. The third position was considered as a 4-fold degenerate site, where only synonymous mutations could appear.

Two population sizes were simulated. One of 270,000 individuals representing large mainland population sizes, and another of 110,000 individuals representing smaller island populations. These values were estimated from the synonymous nucleotide diversity (P_s) estimated for insular and mainland species in our dataset (see part *Polymorphism analyses* from Materials) which allow the estimation of the effective population size as $N_e = P_s / 4 * \mu$ (with $\mu = 4.6e-9$ corresponding to mutation rate estimated for the collared flycatcher in Smeds et al. 2016). Simulations were replicated 15 times, by sampling the genome of 15 individuals after 100,000 generations, of which 10,000 generations corresponded to a burning phase. Nucleotide diversity of non-degenerated and 4-fold degenerate sites were estimated by arithmetic averaging the divergences of all sequence pairs (Tajima 1989).

Next, we simulated the effect of population size variation on P_n/P_s under two types of balancing selection, namely frequency-dependence and overdominance (Figure S4, S5). In both cases, we simulated 10 population sizes for both island and mainland populations (starting respectively from 1100 and 2700 individuals and doubled between each population size step), 10 replicas were made for each population size. We use the same parameters as before, except for the non-synonymous mutation, where for i) frequency dependence, mutations were initially set a fitness effect of 0.01 and then re-estimated at each generation according to the mutation frequency in the population, and ii) for overdominance, mutations sets a fitness effect of 0.01 and a dominance coefficient of 1.5.

Finally, we tested the effect of selection coefficient variation. To do so, we fixed population size at 110,000 individuals and made the fitness vary from 0 to 0.045 for frequency dependence mutation (Figure S6A), and the dominance coefficient (h) from 1.0 to 1.8 for overdominance mutations (Figure S6B).

Supplementary figures and tables:

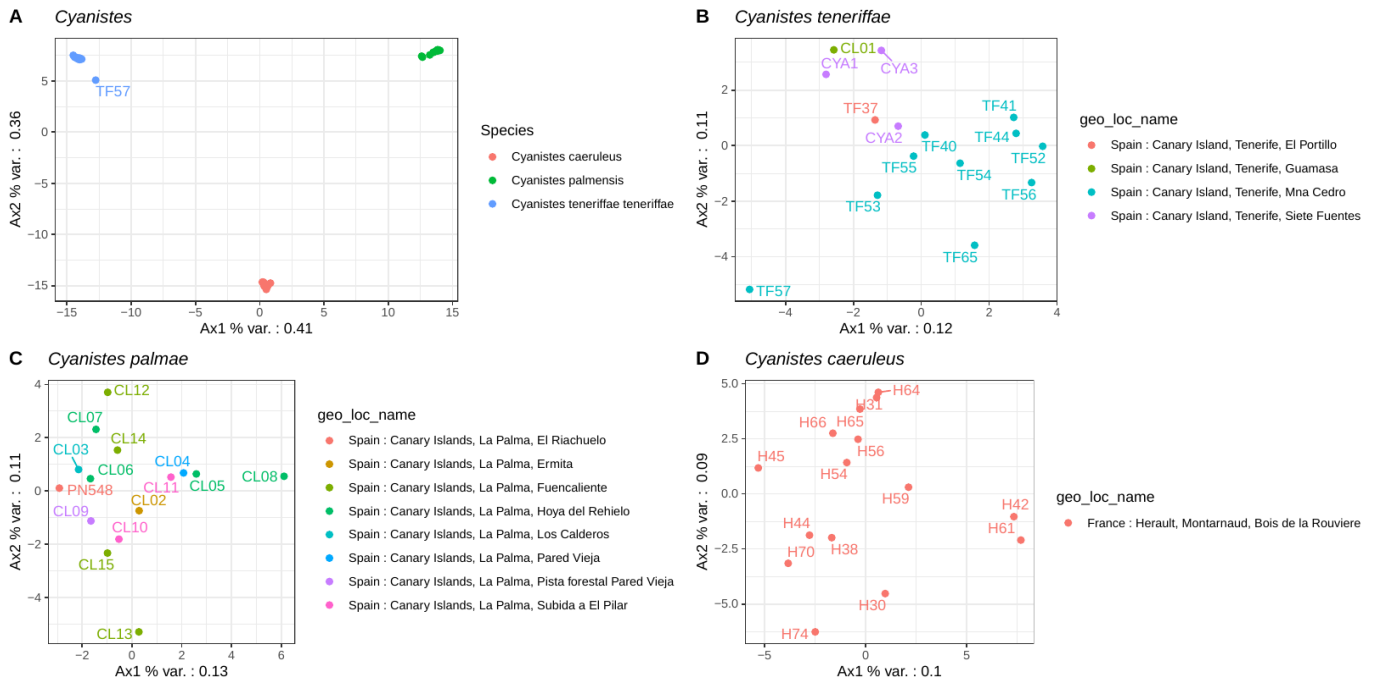


Figure S2: Principal Component Analysis PCA using alleles frequencies of the control genes. A) All *Cyanistes* species; B) *Cyanistes teneriffae*; C) *Cyanistes palmensis*; D) *Cyanistes caeruleus*. Colors indicate species or locality.

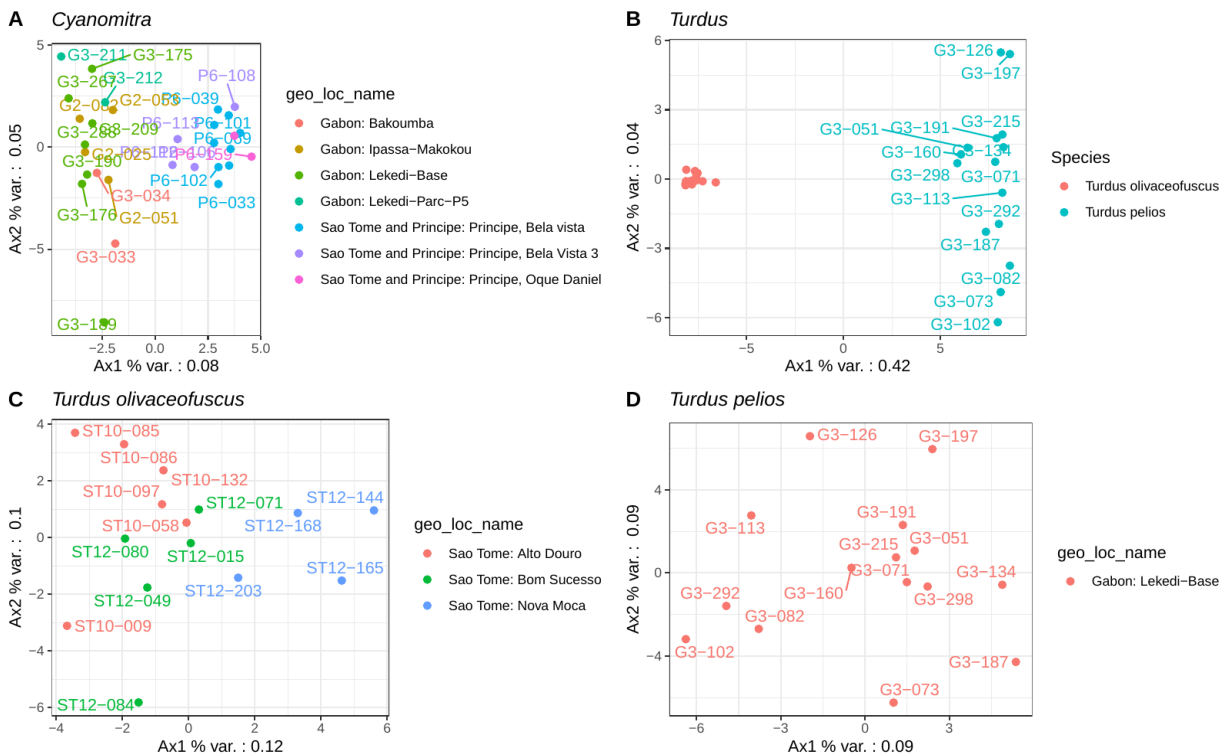


Figure S3: Principal Component Analysis PCA using alleles frequencies of the control genes. A) *Cyanomitra olivacea*; B) All *Turdus* species; C) *Turdus olivaceofuscus*; D) *Turdus pelios*. Colors indicate species or locality.

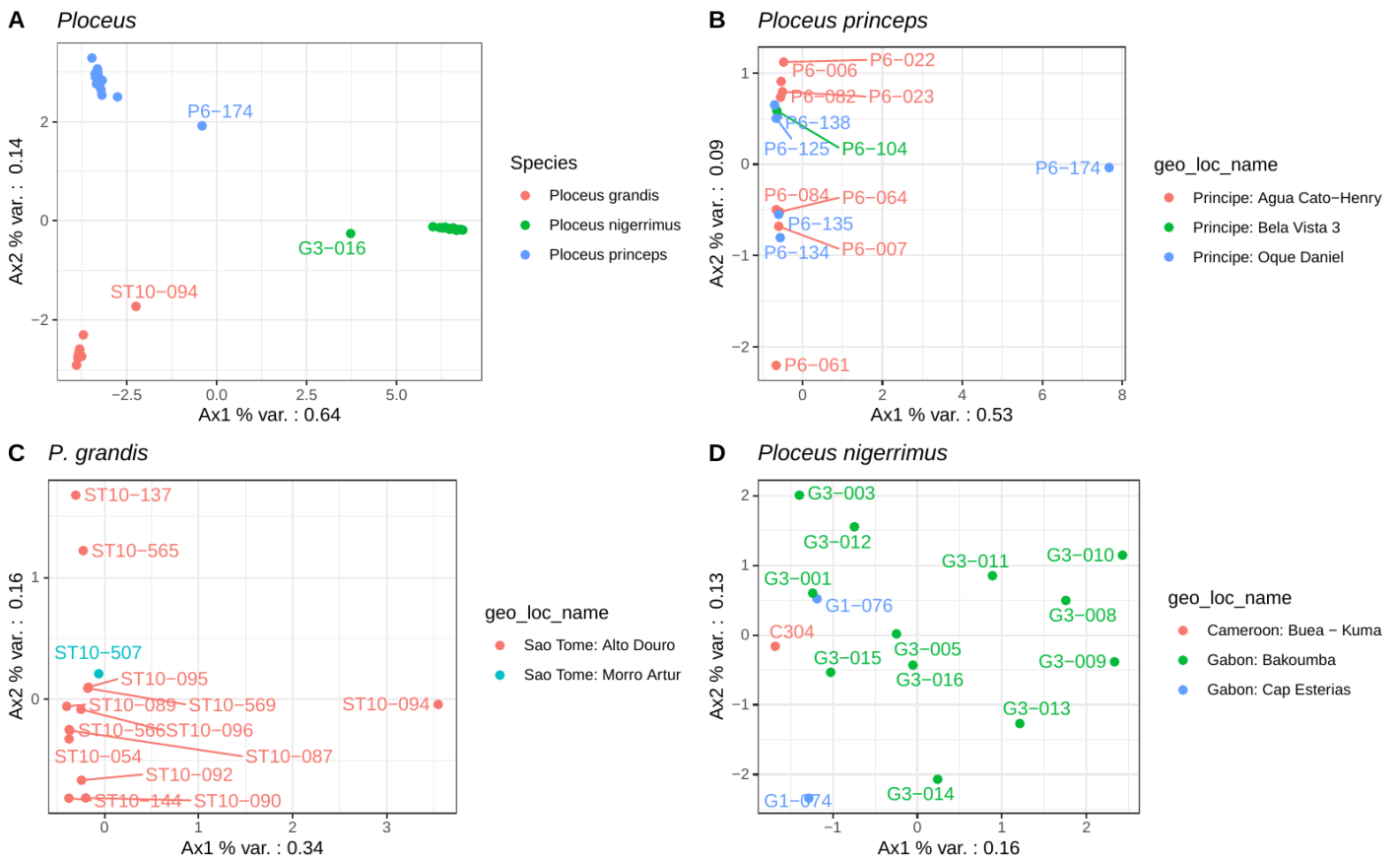


Figure S4: Principal Component Analysis PCA using alleles frequencies of the control genes. A) All *Ploceus* species; B) *Ploceus princeps*; C) *Ploceus grandis*; D) *Ploceus nigerrimus*. Colors indicate species or locality.

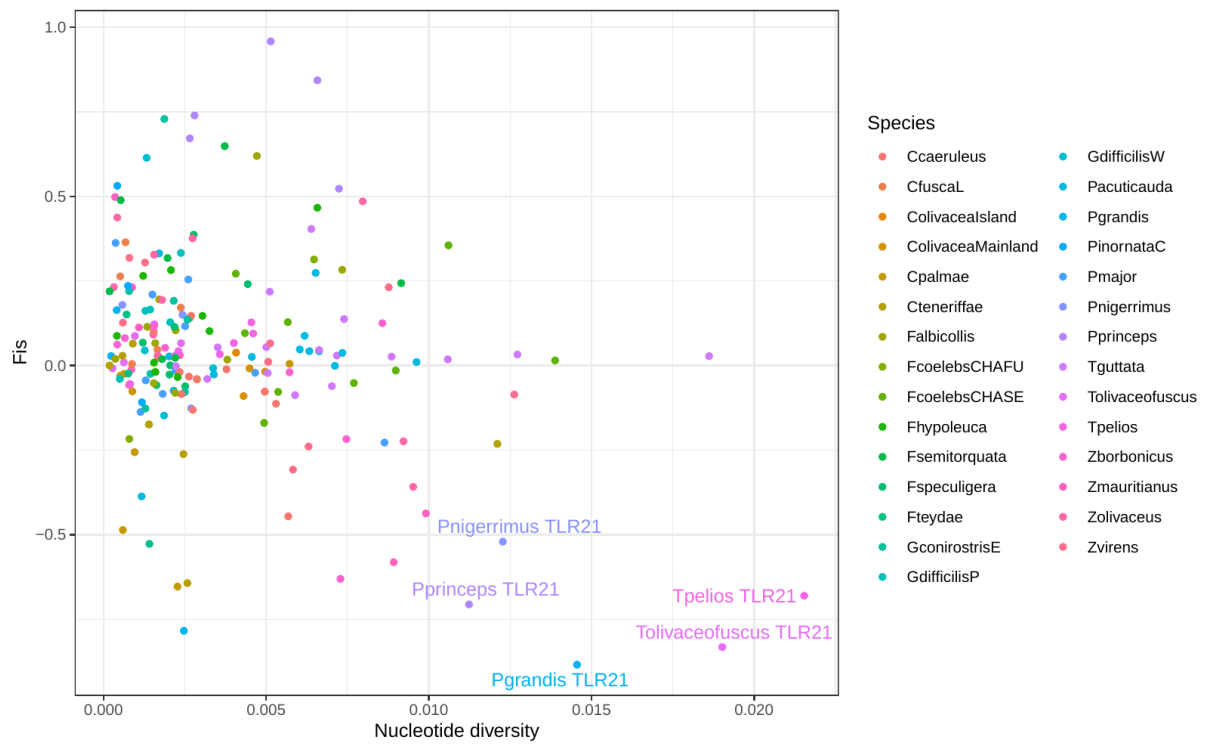


Figure S5: Deviation of heterozygosity from Hardy-Weinberg proportion (F_{is}) according to nucleotide diversity (π). A negative F_{is} indicates more heterozygous individuals than expected.

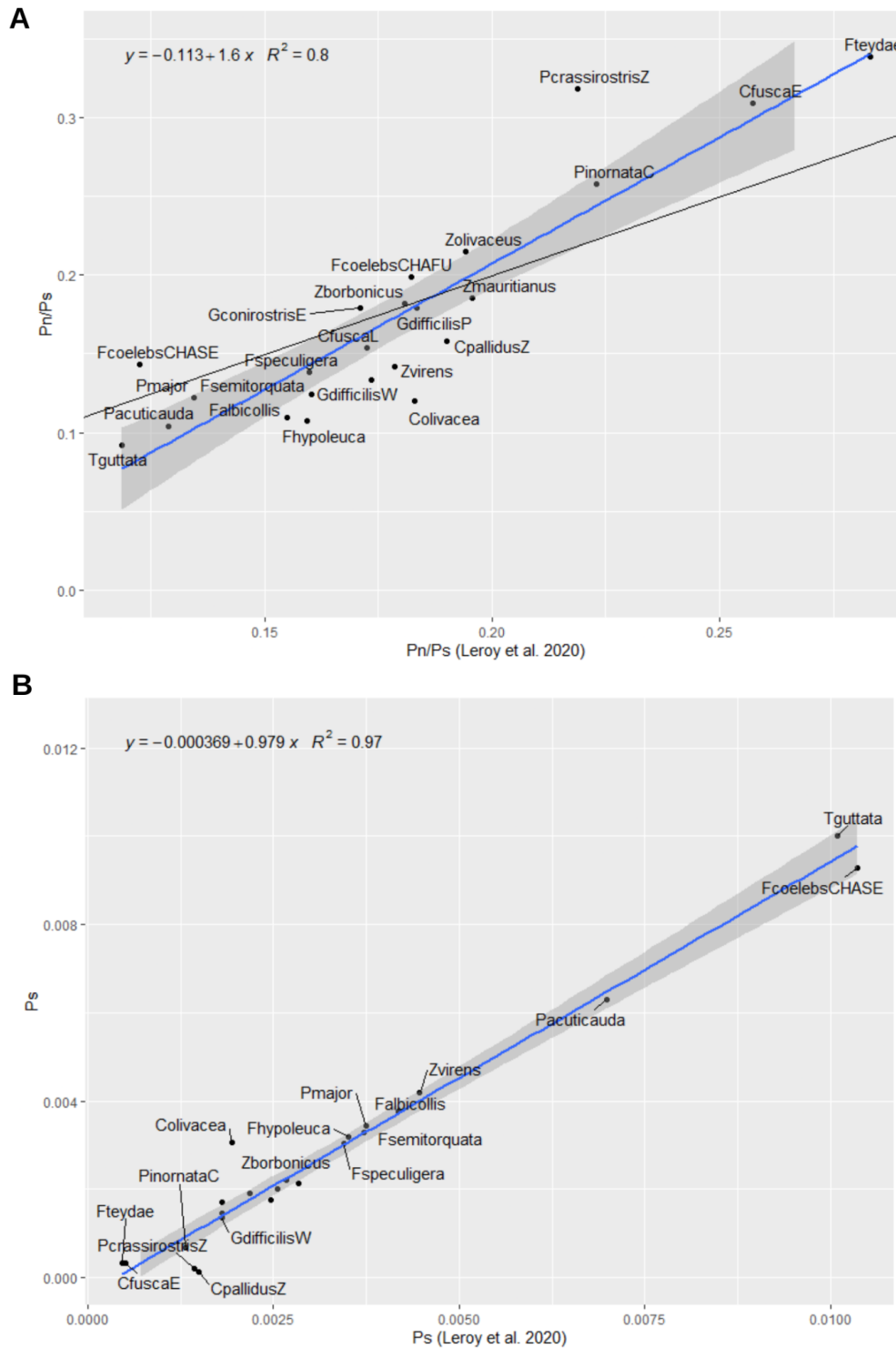


Figure S6: Correlation between Pn/Ps (a) and Ps (b) calculated on the control genes in this study's dataset and those calculated by Leroy et al. (2021). Ps: synonymous polymorphism
Pn: non-synonymous polymorphism.

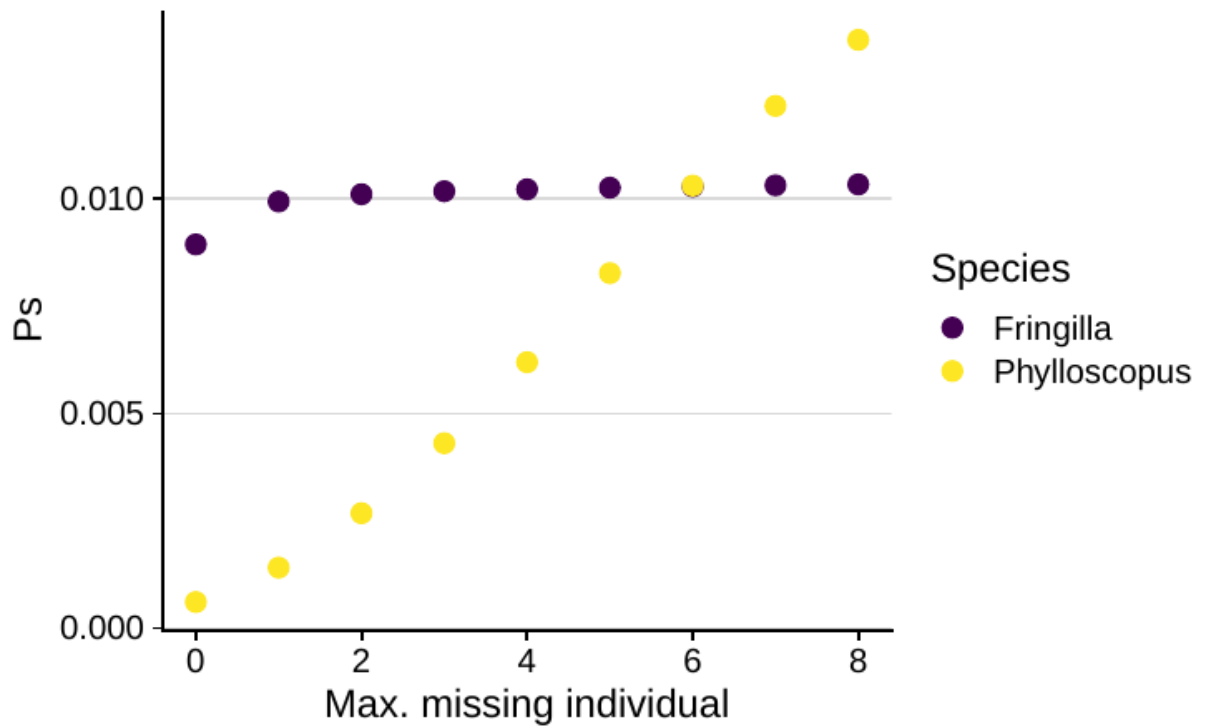


Figure S7: Relationship between the maximum number of missing individuals allowed and synonymous nucleotide diversity (P_s) in *Phylloscopus trochilus* and *Fringilla coelebs*.

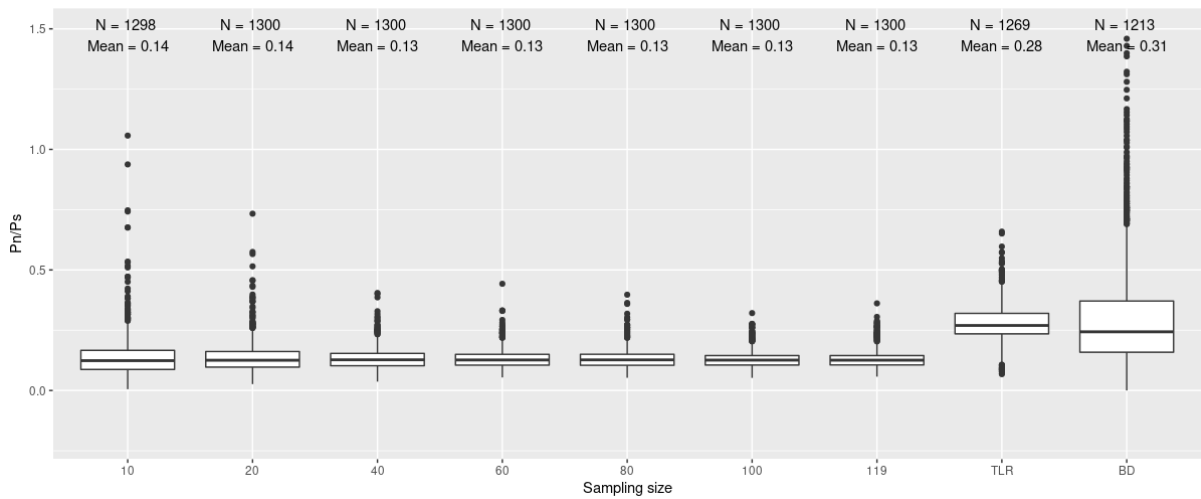


Figure S8: Effect of sub-sampling size on PN/PS. Bootstrap of 100. For BD sampling of 9 genes, for TLR sampling of 10 genes.

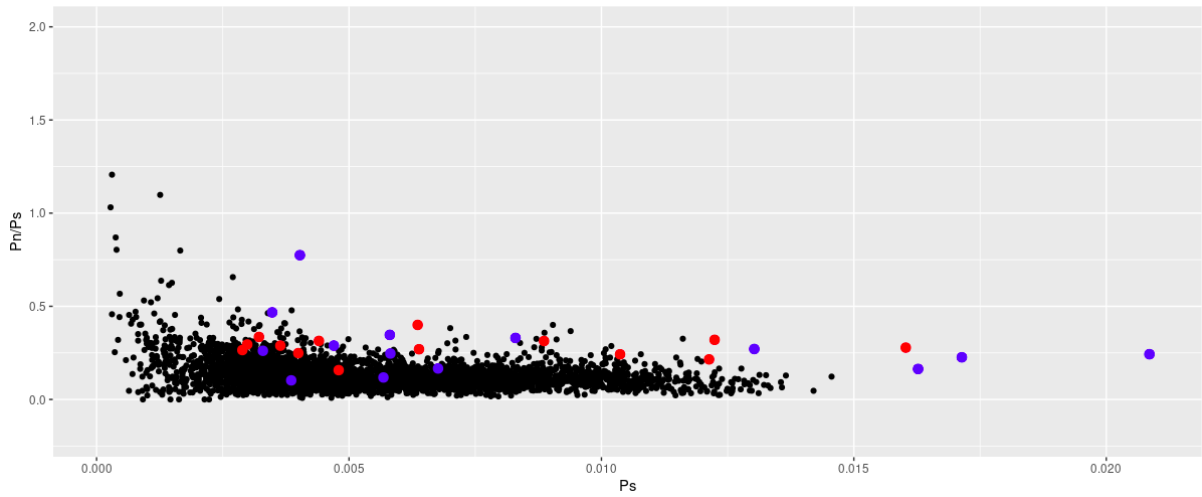


Figure S9: Effect of sub-sampling of control genes, here each point corresponds to the values of a mainland species calculated on 10 randomly selected control genes. In blue the BD, in red the TLR gene.

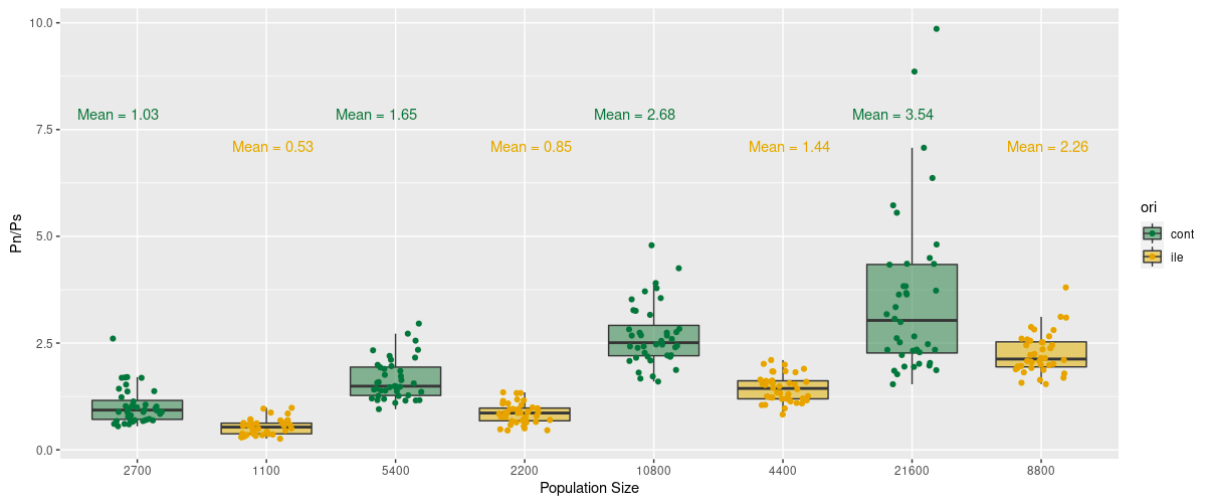


Figure S10: Boxplot of P_n/P_s according to population size for simulated sequences under **overdominance** with SLiM. Dominance coefficient is fixed at 1.5. Each modality is replicated 10 times.

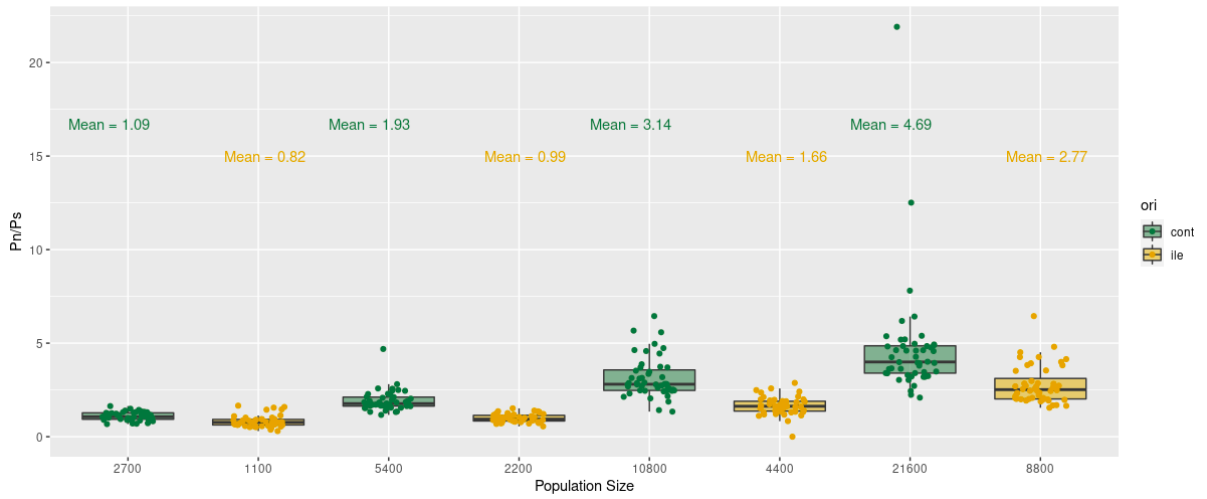


Figure S11: Boxplot of P_n/P_s according to population size for simulated sequences under frequency dependence with SLiM. Each modality is replicated 10 times.

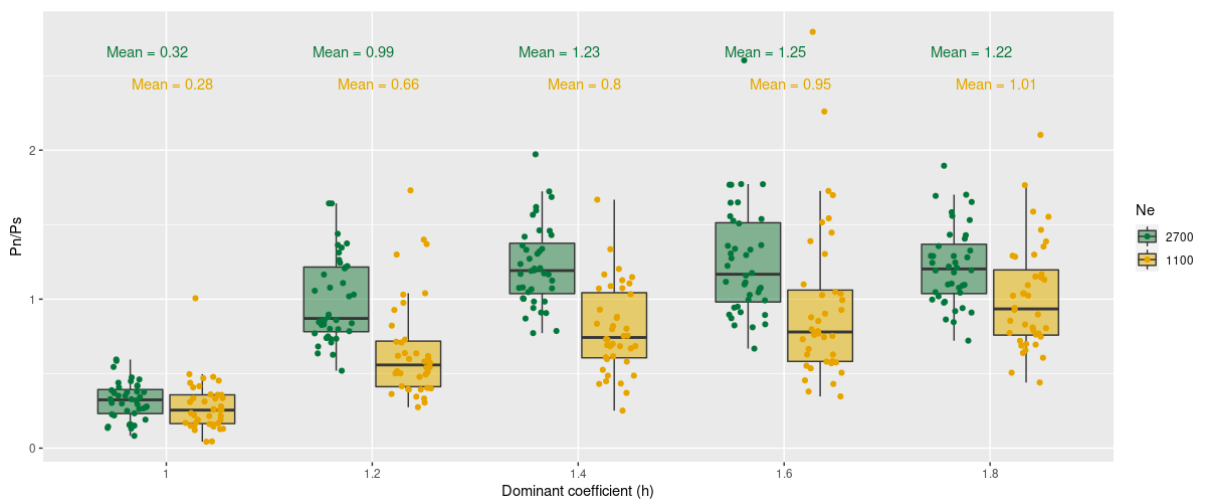
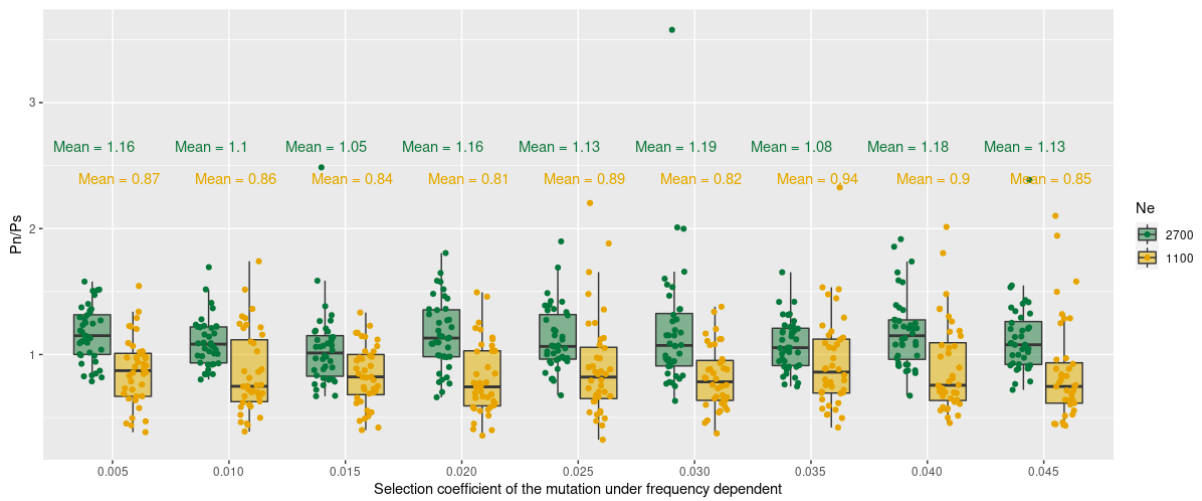


Figure S12: Boxplot of Pn/Ps according to a) initial selection coefficient of the mutation under frequency dependence b) dominance coefficient (h) for simulated sequences under overdominance with SLiM. Each modality is replicated 10 times.

Table S1: Model selection of all gene categories using a reduced number of families (we grouped Turdidae within Muscicapidae, Nectariniidae, and Estrildidae within Ploceidae and Fringillidae within Thraupidae).

<https://figshare.com/s/ab7004cc2f4415b4058f>

Table S2: Summary of the linear model and PGLS model using phylogeny from Figure 1 and the $\Delta Pn/Ps$ (difference between the Pn/Ps of immune genes and control genes).

<https://figshare.com/s/ab7004cc2f4415b4058f>

Table S3: Table with information regarding the samples newly-sequenced in this study.

<https://figshare.com/s/ab7004cc2f4415b4058f>

Table S4: Table with information regarding the sample obtained from Leroy et al. 2021.

<https://figshare.com/s/ab7004cc2f4415b4058f>

Table S5 to S14: Model selection by AICc criterion and ANOVA test. Summary of the best models.

<https://figshare.com/s/ab7004cc2f4415b4058f>

Table S15 to S24:

<https://figshare.com/s/ab7004cc2f4415b4058f>

References

- Haller BC, Messer PW. 2017. SLiM 2: Flexible, interactive forward genetic simulations. *Mol. Biol. Evol.* 34:230–240.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676.
- Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX. *BMC Bioinformatics* 12:1–9.
- Mistry J, Finn RD, Eddy SR, Bateman A, Punta M. 2013. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. *Nucleic Acids Res.* 41:e121–e121.
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16:276–277.
- Rousselle M, Simion P, Tilak M-K, Figuet E, Nabholz B, Galtier N. 2020. Is adaptation limited by mutation? A timescale-dependent effect of genetic diversity on the adaptive substitution rate in animals. *PLoS Genet.* 16:e1008668.
- Siewert KM, Voight BF. 2020. BetaScan2: Standardized Statistics to Detect Balancing Selection Utilizing Substitution Data. *Genome Biol. Evol.* 12:3873–3877.
- Simion P, Belkhir K, François C, Veyssier J, Rink JC, Manuel M, Philippe H, Telford MJ.

2018. A software tool ‘CroCo’ detects pervasive cross-species contamination in next generation sequencing data. *BMC Biol.* 16:1–9.
- Smeds L, Qvarnstrom A, Ellegren H. 2016. Direct estimate of the rate of germline mutation in a bird. *Genome Res.*:gr-204669.
- Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Velová H, Gutowska-Ding MW, Burt DW, Vinkler M, Yeager M. 2018. Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. *Mol. Biol. Evol.*